

RESEARCH

Open Access



Transcriptome-scale analysis of functional alternative back-splicing events in colorectal cancer

Qianqian Ning^{1,2,3†}, Qian Jin^{1,2,3†}, Lei Zhao^{4†}, Yudi Wang¹, Jie Wang¹, Lili Yang¹, Ye Han⁵, Qiaoming Zhi⁵, Junnian Zheng^{2,3*}, Feifei Chen^{1,2,3*} and Dong Dong^{1,2,3*}

Abstract

Background Circular RNAs (circRNAs) are a class of non-polyadenylated RNAs generated from back-splicing of genes. Multiple circRNAs can be generated at a single gene locus through alternative back-splicing events (ABS), sharing the same 5' or 3' back-splice site. To date, how prevalent ABS events are and how they are participated in carcinogenesis of human colorectal cancer (CRC) remains unexplored.

Methods To explore the functional roles of ABS events in CRC carcinogenesis, we analyzed ribosomal RNA-depleted transcriptome sequencing data of 176 CRC samples and characterized the landscape of ABS events in CRC. CRC cancer-related ABS events were identified by comparing paired CRC tumor tissues and adjacent normal tissues. Then, univariate and multivariate Cox regression was used to find prognostic ABS events. Moreover, in vitro and in vivo assays were used to exploring the functional roles of *circXPO1-1* and *circXPO1-2* in CRC.

Results We totally identified 19,611 high confidence circRNAs in CRC, among which 17,874 (91.1%) of circRNAs were found recurrently. The number of ABS circRNAs accounted for 68.8% of all identified high confidence circRNAs, which suggested that ABS events are prevalent in CRC transcriptome. Particularly, 552 ABS circRNAs were found to be aberrantly expressed between paired CRC tumor tissues and adjacent normal tissues, and their parent genes are closely associated with cancer-related hallmarks. In addition, 13 differential ABS circRNAs were identified to be associated with CRC patient survival and could act as independent prognostic indicators. Furthermore, we identified two ABS circRNAs of *XPO1* gene (*circXPO1-1* and *circXPO1-2*). The result showed that overexpression of *circXPO1-2* inhibited CRC cell proliferation, migration, and invasion in vitro and in vivo, whereas *circXPO1-1* is not, indicating that the circularization isoforms of *XPO1* gene have different functions in CRC.

Conclusions In conclusion, our work provides the landscape of ABS events in CRC transcriptome and the close association of ABS circRNAs with tumorigenesis offers a new set of targets with potential clinical benefit.

Keywords Circular RNA, Alternative back-splicing event, Colorectal cancer, CircXPO1

[†]Qianqian Ning, Qian Jin and Lei Zhao have contribute equally to this study.

*Correspondence:

Junnian Zheng

jnzhen@xzhmu.edu.cn

Feifei Chen

chenfeifei@xzhmu.edu.cn

Dong Dong

ddong@xzhmu.edu.cn

Full list of author information is available at the end of the article



Background

Circular RNAs (circRNAs) are a class of endogenous and covalently closed RNAs formed by non-sequential back-splicing from pre-mRNAs [1, 2]. Thanks to the recent advances in next-generation sequencing technologies, circRNAs have become widely identified in human samples, which largely expanded the transcriptome complexity [3–7]. CircRNAs are mainly produced from spliceosome-mediated pre-mRNA splicing wherein a downstream 5′ splice site (splice donor) is joined to an upstream 3′ splice site (splice acceptor). In general, circRNAs are more stable than the linear RNA isoforms derived from the same host gene [8, 9]. Increasing researches showed that circRNAs play critical roles in different biological and pathophysiological conditions, through modulating transcription and splicing, regulating mRNA stability and translation, or serving as templates for translation etc. [2, 10]. The functions of circRNAs in regulating cell proliferation, epithelial-mesenchymal transition (EMT), tumor metastasis, and drug resistance provide the rationale for anticancer therapy [11–15].

Previous works have documented that a single gene locus can produce diverse circRNA isoforms through alternative back-splicing [6, 16]. It is a circularization mechanism associated with the competition of inverted complementary sequences cross introns that bracket the circRNA-forming exons. In theory, there are two types of ABS events, alternative 3 back-splicing (A3BS) and alternative 5 back-splicing (A5BS) events. In the A3BS event, A3BS circRNAs share the same downstream 5′ back-splice site, whereas circRNAs have the same upstream 3′ back-splice site in A5BS events [16]. Thus, circRNA repertoires are largely expanded though ABS events. However, few works have comprehensively analyzed the ABS events and their detailed functions are still largely unknown, especially in human cancers.

In this work, we systematically profiled and characterized the landscape of ABS events in CRC based on rRNA-depleted RNA-seq data. The result showed that ABS events are prevalent in human CRC transcriptome, and their parent genes are closely associated with cancer-related hallmarks. We identified several CRC-related ABS circRNAs and investigated their associations with CRC clinical outcomes. Furthermore, we identified two ABS circRNA isoforms of XPO1 gene, *circXPO1-1* and *circXPO1-2*, which have different expression and function patterns. Taken together, our work identified the widespread CRC-related ABS circRNAs and provided a valuable resource for depicting the biogenesis and progression of CRC.

Methods

Human subjects and data collection

A total of 78 tumor human CRC and corresponding benign adjacent samples were acquired from the Department of General Surgery, the First Affiliated Hospital of Soochow University (Suzhou, China) with informed consent and this study was conducted following approval by the ethical committee of the First Affiliated Hospital of Soochow University. Subsequently, we collected three additional datasets from the NCBI SRA database (SRP158734, SRP291797) and National Genomics Data Center (PRJCA001113), including 88 tissue samples for further analysis. The details of the datasets used in this study were listed in Additional file 2: Table S1. In total, 126 primary tumors and 50 adjacent normal tissue samples were curated, and all corresponding RNA-seq data were generated from the Illumina sequencing platform (rRNA-depleted RNA-seq with PE150 library).

RNA isolation and sequencing

Total RNA of 88 CRC patient samples were extracted using RNeasy Mini Kit (Qiagen, Germany) in accordance with the manufacturer's protocol. RNA integrity was verified using Agilent Bio-analyzer 2100 (Agilent Tech, CA, USA). Complementary DNA was synthesized from 2 μg of total RNA by utilizing Superscript III and random primers. For each sample, ribosomal depleted library was constructed using Illumina's TruSeq Total RNA Library Prep Kit with Ribo-Zero. End-repair and adaptor ligation were performed in accordance with the manufacturer's protocols. Library quality was approved by assaying libraries on an Agilent 2100 Bioanalyzer of product size and concentration. Then, RNA-seq data were generated based on Illumina Nova-Seq 6000 platform.

Identification of circRNAs and ABS circRNAs

First, the quality of raw RNA-seq data was evaluated using Fastp v0.20.1 software [17] (<https://github.com/OpenGene/fastp>), and all obtained clean reads were qualified for downstream analysis. Second, we used the CIRIquant pipeline [3] to identify circRNAs based on rRNA-depleted RNA-seq data. CIRIquant provides more accurate expression values for circRNAs with significantly reduced false discovery rate by constructing a pseudo-circular reference for re-alignment of RNA-seq reads. RNA-seq reads for each sample were firstly mapped to the hg38 human reference genome with the GENCODE gene annotation v34 by Hisat2 [18] (Version 2.2.0). Then, the unmapped reads were re-aligned to the pseudo reference of circular sequences, which is generated by repeating whole length of back-spliced region twice using Hisat2. The reads aligned to the back-spliced

junction sites were used to calculate circRNA junction ratio and evaluate differentially expressed circRNAs. Third, we further used CIRCexplorer2 [6] to determine the ABS circRNAs and ABS events. CIRCexplorer2 is a system which comprehensively deciphers the alternative back-splicing/splicing pattern of circRNAs.

An ABS event is defined as a group of circRNAs that share the same 5' back-splice site or 3' back-splice site according to the genomic coordinates of the circRNAs. We first defined high-confidence circRNAs as the circRNAs with ≥ 0.1 RPM in at least one sample. Then a high-confidence ABS event was defined as a group of two or more high-confidence circRNAs that sharing a same back-splice site. In this study, we mainly focused on those high-confidence ABS circRNAs and high-confidence ABS events.

Alu repeats in introns

The annotation file of *Alu* repeats was downloaded from the UCSC Browser RepeatMasker track. Positions of *Alu* repeats were intersected with the flanking intron regions of circRNAs or random introns.

Identifying CRC tumor-specific ABS circRNAs

To systematically evaluate ABS circRNAs associated with CRC, we calculated the Percent Circularized-site Usage (PCU) value for each ABS circRNA. PCU was defined as the ratio of the number of back-splice junction reads of each circRNA to the total number of back-splice junction reads of all ABS circRNAs sharing the same back-splice site. Then, we computed the difference of PCU values (Δ PCU) for each high-confidence ABS circRNA between paired CRC tumor and normal tissues. The ABS circRNAs with $|\Delta$ PCU > 0.1 and P -value < 0.05 (Student's t test) were considered as CRC tumor-specific.

Construction splicing factor and ABS circRNA dysregulation network

A catalogue of 67 annotated splicing factors (SFs) was retrieved from the previous report [19]. Spearman correlation coefficients (ρ) between the expression level of SFs and the PCU values of ABS circRNAs were calculated. Then, SFs and ABS circRNAs with absolute ρ values > 0.5 and P -value < 0.05 were considered to be significantly correlated. The dysregulation network was built and visualized by Cytoscape (version 3.6.1).

Functional enrichment analysis

The parent genes were extracted from corresponding circRNAs to detect their functional implications. A hallmark gene set was downloaded from the MsigDB database [20], which summarizes and represents specific well-defined biological states or processes. These 50

biological processes are closely related to the hallmarks of cancer containing a total of 7,321 genes. The enrichment analysis was performed using the clusterProfiler package (version 3.12.0) [21] using the threshold of P -value < 0.05 . For Gene Set Enrichment Analysis (GSEA), all parent genes of ABS circRNAs with ranked Δ PCU values were used. Additionally, oncogenes and tumor suppressor gene lists were obtained from the NCG database [22].

Survival analysis

The survival analysis was performed based on the clinical information of our cohort. Univariate Cox regression followed by multivariate analysis was performed based on overall survival to determine independent prognostic factors. For each DE ABS circRNA, CRC patients were divided into two groups based on the PCU values, and the two-category was modeled as continuous variables to obtain a hazard ratio (HR). Kaplan–Meier analysis with a log-rank test was used to compare patients' survival between subgroups.

Cell culture, RNA isolation, gDNA extraction and qRT-PCR

Human CRC cell lines (SW620 and LoVo) were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). All these cells were cultured in DMEM (Dulbecco's modified Eagle's medium) with 10% FBS (fetal bovine serum), 50 U/mL penicillin and 0.1 mg/mL streptomycin at 37 °C with 5% CO₂. All cells were negative for mycoplasma contamination.

Total RNA of cell lines and tissues was extracted using RNAPure Tissue&Cell Kit (DNase I) (CoWin Biotech Co., Ltd.) according to the manufacturer's protocol. Complementary DNA was synthesized using 500 ng total RNA as template in a final volume of 10 μ L using the PrimeScript RT reagent kit (TaKaRa, Japan). Genomic DNA (gDNA) was extracted from cultured cells using the FastPure Cell/Tissue DNA Isolation Mini Kit (Nanjing Vazyme Biotech Co., Ltd.) according to the manufacturer's protocol. The qRT-PCR was conducted using 0.25 μ L of the cDNA solution in a final volume of 20 μ L with SYBR Premix Ex Taq (Transgene) in 7600 Real-Time PCR System (Applied Biosystems). GAPDH was used as an internal control. All primers used in this study were listed in Additional file 2: Table S10.

RNA isolation of nuclear and cytoplasmic fractions and RNase R treatment

The subcellular localization of *circXPO1-1* and *circXPO1-2* were investigated using a Nuclear and Cytoplasmic Protein Extraction Kit (Beyotime Biotechnology) plus with RNase inhibitor, Trizol reagent, a modified protocol for RNA isolation. RNase R treatment was executed

at 37 °C with 4U/μg of RNase R (Epicentre Biotechnologies, Madison, WI, USA) for 30 min.

FISH assay

Cy3-labelled *circXPO1-1* and Cy3-labelled *circXPO1-2* probes were designed and synthesized. A Fluorescent In Situ Hybridization Kit (Guangzhou RiboBio Co., Ltd.) was used to detect the probe signals in LoVo cells according to the manufacturer's protocol. Cell nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI). All images were photographed under the LSM880 confocal microscope system (Carl Zeiss).

Plasmids and cell transfection

For *circXPO1-1* and *circXPO1-2* overexpression, the sequence of exon 3, 4, 5 and exon 2, 3, 4, 5 from transcript ENST00000680228 were synthesized and cloned into GV727 vector (Shanghai Genechem Co., LTD.), respectively. Then, the GV727, GV727-*circXPO1-1* and GV727-*circXPO1-2* vectors were transfected into SW620 and LoVo cells using Lipofectamine 2000 (Invitrogen, USA).

Cell proliferation and colony formation assays

The effects of *circXPO1-1* and *circXPO1-2* on the cell proliferation were tested with Cell Counting Kit-8 (Dojindo Laboratories, Kumamoto, Japan). Cells were seeded at a density of 1000 cells/well in 96-well plates. After added 10 μL CCK-8 solution, cells were incubated for 2 h at 37 °C. Then the absorbance at 450 nm was measured. Each measurement was performed in triplicate. For the colony formation assays, cells were harvested and seeded in 6-well plates at a density of 600 cells per well. After 14 days, colonies were fixed with 4% neutral formalin and stained with crystal violet, and then imaged using an IX73 inverted microscope (Olympus, Japan).

Cell migration and invasion assays

To test the effects on migration and invasion of CRC cells, the transwell chambers were used. Eighty thousand LoVo or SW620 cells were suspended in 100 μL DMEM and were seeded in the top chamber with (migration) or without (invasion) Matrigel-coated membrane (pore size, 8 μm) (BD Biosciences, USA). The lower chambers were filled with DMEM containing 20% FBS. After 36 h incubation, cells that did not migrate through the pores were removed with a cotton swab. The remaining cells were fixed, stained with crystal violet and counted in random five fields. All experiments were performed in triplicate.

Animal experiments

Tumorigenesis in nude mice was performed as described. Briefly, five weeks old female nude mice were injected with LoVo cells stably transfected with control or *circXPO1-1/2* over-expression plasmids. Tumor volume was measured with calipers at the site of injection weekly ($V = 0.5 * \text{length} * \text{width}^2$). The tumors were obtained and weighted 4 weeks after the injection. All experimental procedures were approved by the Ethics Committee of Xuzhou medical University.

RNA-seq analysis of overexpression of *circXPO1-2*

Trizol reagent was used to extract total RNAs from LoVo cells transfected with control vector or *circXPO1-2* over-expression vector. Three biological replicates were available for each group. Libraries were constructed using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations and then subjected to Nova-Seq 6000 platform. RNA-seq raw data were cutadapter and filtered by Fastp v0.20.1 [17]. Then, all obtained clean reads were assembled to human genome (hg38) by Hisat2 [18] and gene expression level was evaluated by stringtie v2.1.2 [23]. The difference in gene expression between *circXPO1-2* over-expression and control cells was assessed by edgeR [24]. P -value < 0.05 was considered to be statistically significant. GSEA analysis was performed based on the genes ranked according to their difference between above two groups.

Statistical analysis

All statistical analysis and figure plotting in this study were performed using R software (version 3.6.1). Additionally, all statistical tests with a P -value < 0.05 were considered to be significant, unless indicated otherwise mentioned.

Results

The characteristics of circRNAs in CRC transcriptome

At first, we performed ribosomal RNA-depleted RNA-seq in 88 CRC samples, including 78 primary CRC tumors and 10 corresponding adjacent benign samples. To systematically characterize the circRNA landscape in CRC, we also aggregated ribosomal RNA-depleted RNA-seq from three additional CRC cohorts [25–27]. Totally, 126 primary CRC tumors and 50 adjacent benign samples were curated in this study (Fig. 1a, Additional file 2: Table S1). All data were generated by Illumina next sequencing platform (PE150 library) and the raw data were processed through our in-house bioinformatic pipeline. After removing low-quality reads, the obtained clean reads were qualified for downstream analysis, with

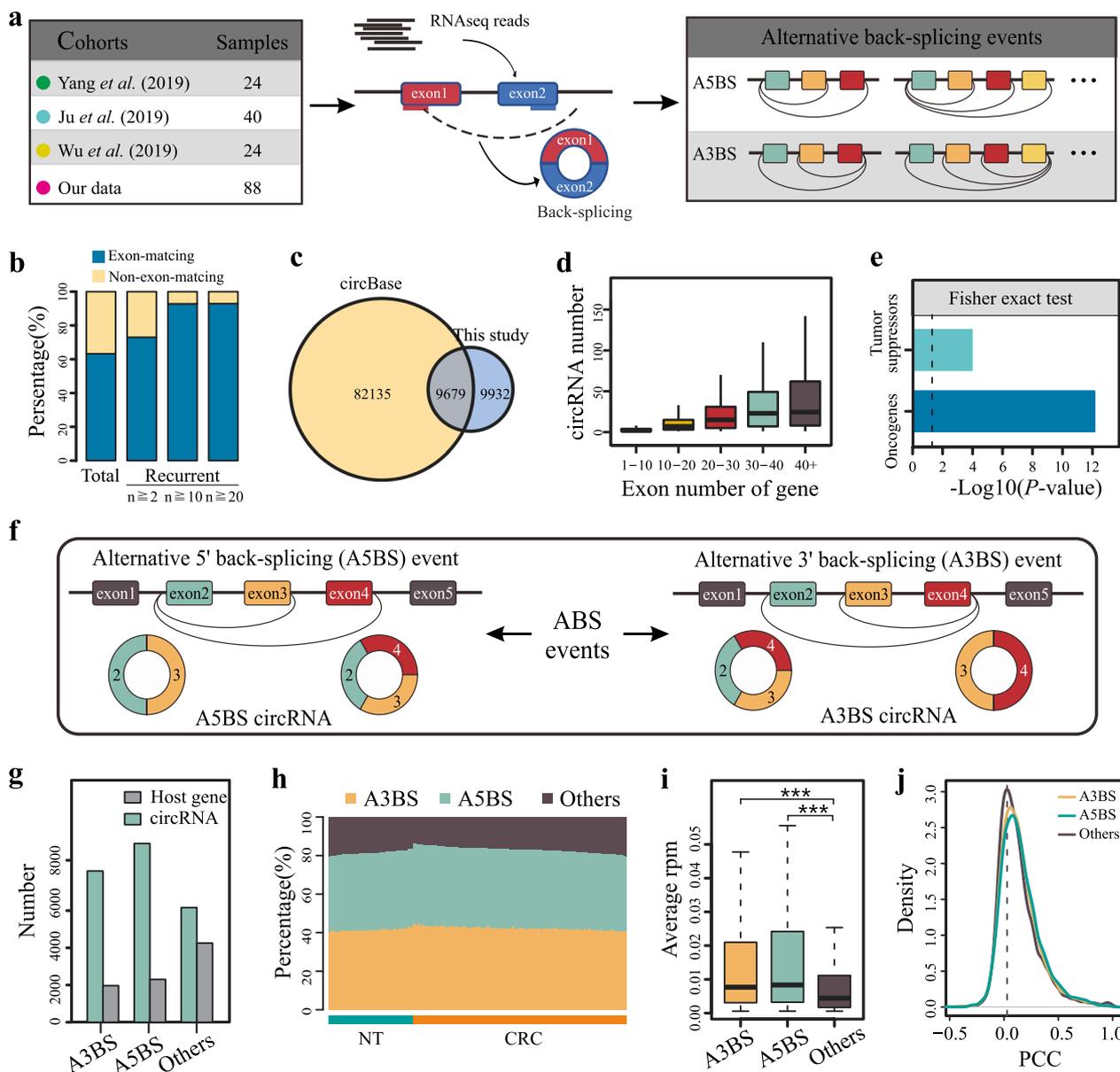


Fig. 1 General characteristics of circRNAs and ABS circRNAs in human CRC tissue transcriptomes. **a** Overview of ABS circRNAs profiling identified by rRNA-depleted RNA-seq. **b** Numbers of unique and recurrent circRNAs in all CRC samples. Unique circRNAs: identified in only one sample. Recurrent circRNAs: identified in at least two samples. Exon-matching: annotation of circulated exon exactly matched with linear counterpart. Non-exon-matching: annotation of circulated exon did not matched with linear counterpart. **c** Venn diagram demonstrating the intersection set of identified high confidence circRNAs and circBase-annotated circRNAs. **d** The number of circRNAs proportionally increases with the exon number of linear counterparts. **e** Oncogenes and tumor suppressor genes with more circRNAs (Fisher's exact test, P -value < 0.05). Oncogenes and tumor suppressor gene lists are obtained from NCG database. **f** Schematic diagrams of two types of alternative back-splicing. A3BS: circRNAs sharing 3' back-splice site. A5BS: circRNAs sharing 5' back-splice site. **g** Bar plot showing the number of ABS events and parent genes from CRC transcriptome according to the ABS event types. **h** Proportions of different ABS event types in CRC tumor and adjacent normal samples. **i** Comparison of circRNAs expression levels of different ABS event types. **j** The distribution of Pearson correlation coefficients (PCC) between circRNA abundance and their parent gene expression across all ABS event types

an average of 91.1 million reads per sample (Additional file 2: Table S2, S3). To evaluate the batch effect of collected data, we analyzed the expression distribution of all

expressed transcripts in these four datasets. The similar distribution of expressed transcripts and the strikingly high expression correlation of 553 housekeeping genes

among these four datasets revealed an acceptable batch effect for further analysis (Additional file 1: Fig S1).

Next, we applied CIRIquant software [3] to identify circRNAs. In total, 303,667 circRNAs were identified, of which 63.2% were generated from exon regions. More than 58% of all circRNAs were only found in one sample (Fig. 1b). To obtain a reliable set of circRNAs, we defined circRNAs with ≥ 0.1 RPM in at least one sample as high-confidence, and about half (49.4%) of 19,611 high-confidence circRNAs were annotated in circBase database (Fig. 1c). As expected, the expression levels of newly identified circRNAs were lower than circRNAs annotated in circBase database (Additional file 1: Fig S2a). A subset of circRNAs that could be consistently detected in >80% samples had relatively higher expression abundance (Additional file 1: Fig S2b). We next examined the expression correlation between circRNAs and their parent genes. The result showed that their correlations were relatively very low (the average Pearson Correlation Coefficient, $R = 0.09$). However, the highly expressed circRNAs (detected in >80% samples) have higher expression correlation with their parent genes (Additional file 1: Fig S2c, the average $R = 0.28$).

Moreover, multiple circRNAs can be generated from the same gene locus. The more exons a gene has, the more circRNA it can produce (Fig. 1d). The number of circRNAs generated from oncogenes or tumor suppressor (TS) genes was significantly higher than other genes (Fig. 1e), suggesting that these circRNAs may be involved in tumorigenesis. To further examine differentially expressed circRNAs, the expression difference of circRNAs was compared between CRC tumor and adjacent normal tissue samples (Additional file 1: Fig S2 d, e). Among those 1992 dysregulated circRNAs, the majority circRNAs (> 92%, 1841) were down-regulated in CRC tumor samples. GSEA enrichment analysis was performed on the parent genes of those dysregulated circRNAs. The result showed that they were over-represented in cancer-related categories, such as PI3 K-AKT signaling pathway, Calcium signaling pathway, MAPK signaling pathway, cAMP signaling pathway, Rap1 signaling pathway, cGMP-PKG signaling pathway, Ubiquitin mediated proteolysis (Additional file 1: Fig S2f).

The prevalence of ABS events in CRC samples

CIRCexplorer2 software [6] was employed to identify ABS circRNAs. According to the 5' back-splice sites and 3' back-splice sites of circRNAs annotated by CIRCexplorer2 software, ABS events that shared 5' or 3' back-splice sites (A5BS and A3BS) (Fig. 1f) were identified (see Materials and Methods), respectively. At last, we totally identified 7142 ABS events (containing 13448 high-confidence ABS circRNAs) in CRC samples, which

accounts for 68.8% of all identified circRNAs (Fig. 1g), suggesting that ABS events are widely prevalent in CRC samples. Obviously, there are more A5BS events than A3BS events in CRC samples (3719 vs. 3423 for A5BS and A3BS events, respectively, Chi-squared test P -value = 8.0×10^{-7}), which indicates that A5BS events are more prevalent than A3BS events. Furthermore, we found that 4,286 circRNAs participated in both A5BS and A3BS events, which accounts for 31.9% of all ABS circRNAs. It indicated that A3BS and A5BS events often shared the same circRNAs. The proportion of ABS circRNAs in CRC tumors was higher than adjacent normal tissue samples (Wilcoxon rank sum test P -value = 8.2×10^{-11} , Fig. 1h). Moreover, both A3BS and A5BS circRNAs have higher expression levels than other circRNAs (Wilcoxon rank sum test P -value $< 2.2 \times 10^{-16}$ and $< 2.2 \times 10^{-16}$ for A3BS and A5BS, respectively, Fig. 1i), and both A3BS and A5BS circRNAs have significantly higher expression correlation with their corresponding parent genes (Wilcoxon rank sum test P -value = 7.6×10^{-10} and 5.7×10^{-21} for A3BS and A5BS, respectively, Fig. 1j).

We randomly selected six A3BS circRNAs (*circASHIL-1*, *circASHIL-2*, *circETNK1-1*, *circETNK1-2*, *circPICALM-1* and *circPICALM-2*) and six A5BS circRNAs (*circARHGAP5-1*, *circARHGAP5-2*, *circCDK13-1*, *circCDK13-2*, *circPTK2-1* and *circPTK2-2*) for experimental validation. Sanger sequencing was performed to confirm the alternative back-splice sites of A3BS and A5BS events after RT-PCR with divergent primers. We found that these ABS events exist in several colorectal cancer cell lines (Fig. 2).

To further examine the genomic features of ABS circRNAs, we analyzed the usage of splice sites and the number of exons for each circRNA. First, circRNAs were produced almost invariably through splicing at canonical splice sites and arose equally from all exons, except those exons that flank gene boundaries (Additional file 1: Fig S3a). A3BS circRNAs had no preference in the front regions of genes, whereas A5BS circRNAs were less produced from back regions of genes (Additional file 1: Fig S3a). The result showed that the majority of circRNAs contain multiple exons, most commonly spanned two to four exons (55.8% of all circRNAs, Additional file 1: Fig S3b). Moreover, ABS circRNAs are likely to contain more exons than other circRNAs (Additional file 1: Fig S3b, ANOVA analysis P -value $< 2.2 \times 10^{-16}$), and the length of ABS circRNAs is much longer (Additional file 1: Fig S3c).

To explore the functional implications of ABS events, we download functional circRNAs from CircFunBase database [28], which is a web-accessible database that aims to provide manually curated high-quality functional circRNAs. The percentage of functional circRNAs in ABS circRNAs (791 of 13,448) was significantly higher than

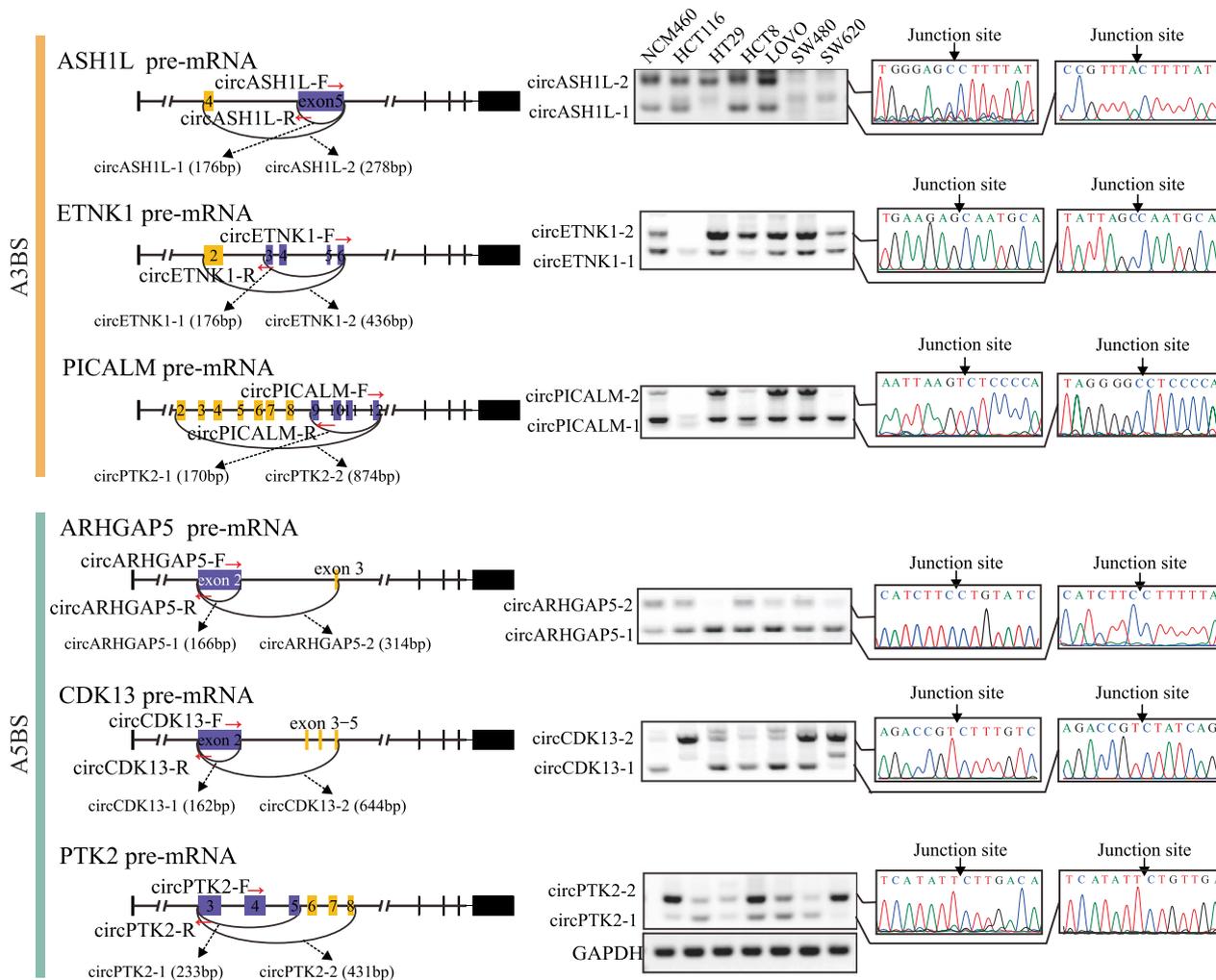


Fig. 2 Experimental validation of ABS circRNAs in CRC cell lines. RT-PCR and Sanger sequencing validation of A3BS circRNAs and A5BS circRNAs. RT-PCR primers of each ABS event were designed at the common region and were divergent. Then the PCR products were subjected to Sanger sequencing after gel extraction

other circRNAs (246 of 6163) (Fisher exact test P -value = 2.38×10^{-8}), suggesting that ABS circRNAs were more preferential to be functional. Additionally, KEGG enrichment analysis was performed based on their parent genes. The result showed that genes harbored A3BS and A5BS events are over-represented in almost similar pathways, like Endocytosis, Regulation of actin cytoskeleton, Proteoglycans in cancer, Protein processing in endoplasmic reticulum and Ubiquitin mediated proteolysis (Additional file 1: Fig S3 d).

The complexity of ABS events in CRC

We next examined the complexity of ABS events in CRC transcriptome. The ABS event complexity is defined by the number of circRNAs belonging to a specific ABS event (Fig. 3a). Therefore, ABS events were divided into two categories based on the complexity (low complexity

category = 2 and high complexity category > 2). As shown in Fig. 3b, most ABS events had low complexity containing only two circRNAs ($n = 4939$, 69.2%), and a subset of ABS events even contain more than five circRNAs ($n = 105$, 1.5%). It seems that the complexity of A5BS events was higher than A3BS events (Fig. 3b, Chi-squared test, P -value < 2.2×10^{-16}). The result showed that ABS events are over-represented in tumor suppressor genes and oncogenes (Fig. 3c). Figure 3d lists the oncogenes/tumor suppressor genes containing ABS events (ABS event complexity > 5). It has been documented that *PICALM* gene plays important role in promoting CRC progression [29]. The highest complexity of ABS events in *PICALM* gene is nine (containing nine A5BS circRNAs, Fig. 3e). The longest circRNA spans from exon 2 to exon 19 (has_circ_0023891), and the shortest one only contains three exons (exon 2–4, has_circ_0023942). The expression of

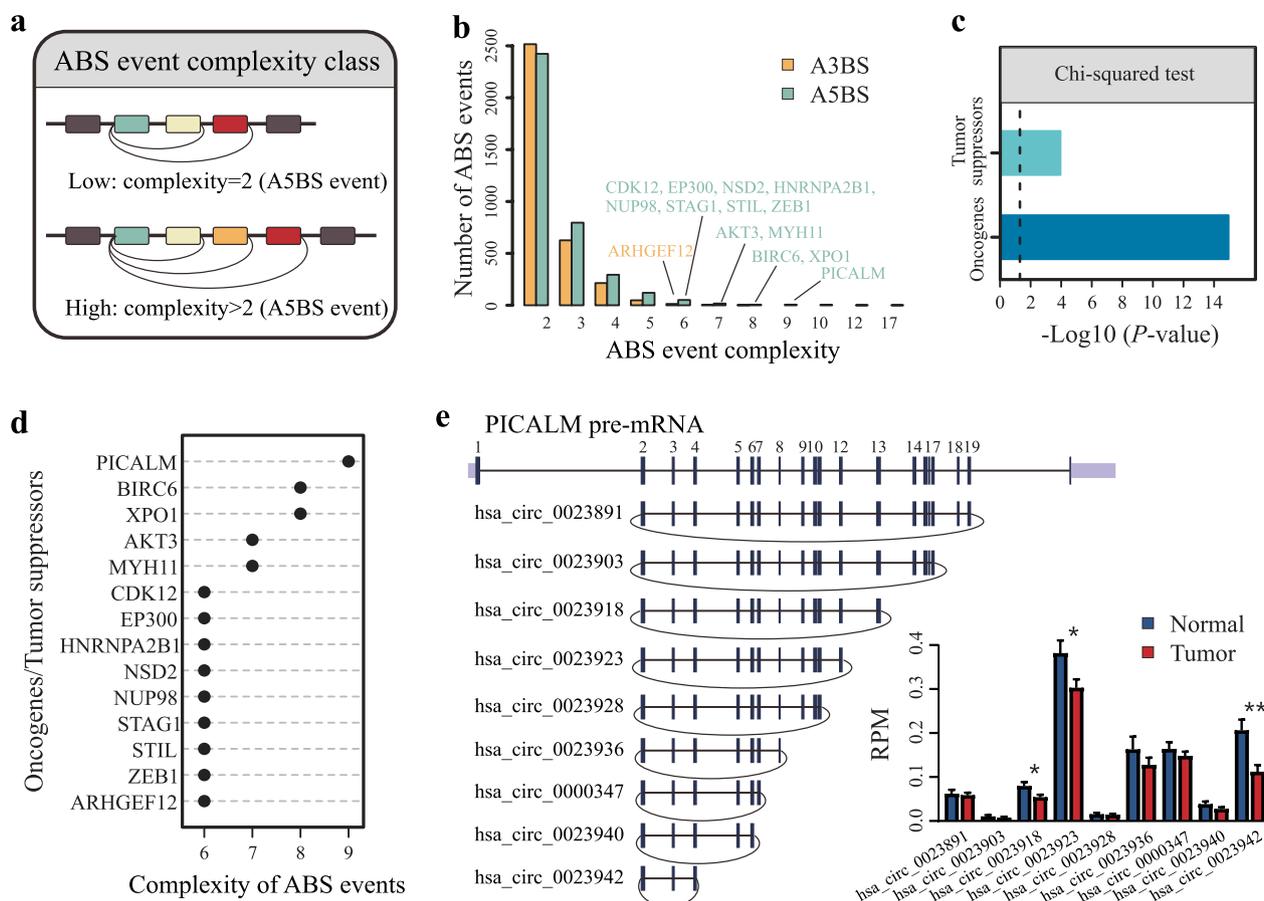


Fig. 3 High complexity of ABS events. **a** Schematic diagram of ABS event complexity. Colored boxes represent exons. Black lines represent introns. Arcs represent back-splicing. ABS events were divided into two groups: low complexity with 2 ABS circRNAs and high complexity with > 2 ABS circRNAs. **b** The distribution of ABS event complexity. The oncogenes or tumor suppressor genes with high complexity ABS events were illustrated. Genes with high complexity A5BS events were highlighted in green and genes with high complexity A3BS events were highlighted in orange. **c** Over-representation of ABS events in tumor suppressor genes and oncogenes. **(d)** Highest complexity of ABS events in selected oncogenes or tumor suppressor genes. **(e)** An A5BS example with high complexity. The top inset shows the structure of an A5BS event with high complexity generated from *PICALM* gene. The bottom inset shows the expression level of each circRNA

these nine A5BS circRNAs was seemingly to be down-regulated in CRC samples (3 of 9 were significant), suggesting that they might function as tumor suppressors in CRC.

It has been reported that reverse complementary sequences greatly influence circRNA production [9, 30, 31]. We herein examined whether *Alu* element count and flanking intron length would have effects on the ABS event complexity or not. The result showed that A5BS events have longer upstream flanking intron length and a higher number of *Alu* elements in upstream regions, whereas A3BS events have longer downstream flanking intron length and a higher number of *Alu* elements in downstream regions (Additional file 1: Fig S4a, b). For A5BS events, circRNAs in high complexity category have longer flanking intron length and higher number of *Alu* elements in upstream regions. For A3BS events, circRNAs in high complexity category have shorter flanking

intron length and a lower number of *Alu* elements in upstream regions; instead, circRNAs in high complexity category have longer flanking intron length and a higher number of *Alu* elements in upstream regions (Additional file 1: Fig S4c, d). This result indicated that genomic features have significant effects on the choice of circRNAs in ABS events.

The properties of CRC specific ABS circRNAs

To compare the difference of ABS events between CRC tumor and adjacent normal tissue samples, we calculated the PCU value for each alternative back-splice site (see Material and Methods). PCU value is an indicator of the proportion of each ABS circRNA in a specific ABS event. The distributions of PCU values for A3BS and A5BS events showed that A3BS events had a higher PCU level than A5BS events in CRC samples (0.48 vs. 0.45,

Wilcoxon rank sum test P -value $< 2.2 \times 10^{-16}$). Both A3BS and A5BS events had higher PCU levels in CRC tumor tissues than adjacent normal tissues (Additional file 1: Fig S5a). We defined differentially expressed ABS (DE ABS) events to be significant with the criterion of $\Delta|PCU| > 0.1$ and P -value < 0.05 between paired CRC tumor and normal tissue samples (Fig. 4a). At last, we obtained a set of 256 DE A3BS circRNAs from 196 genes and 296 DE A5BS circRNAs from 225 genes, among which 42 genes contain both DE A3BS and A5BS circRNAs (Fig. 4b, Additional file 1: Fig S5b, Additional file 2: Table S4, S5). We randomly selected two DE A3BS events nested in *MKLN1* and *PTBP3* genes, and two DE A5BS events in *GDI2* and *RAB11 FIP1* genes. Figure 4c-f presented the PCU value differences of these 4 ABS events between CRC and normal tissue samples.

To further investigate the functional implication of DE ABS circRNAs, we performed GSEA analysis with KEGG and Reactome pathway database. The KEGG results showed that these genes were highly over-represented in cancer-related biological pathways, such as the PI3 K-Akt signaling pathway (P -value = 0.0021), Hippo signaling pathway (P -value = 0.0022), Ras signaling pathway (P -value = 0.0022), MAPK signaling pathway (P -value = 0.0022), Wnt signaling pathway (P -value = 0.0022), mTOR signaling pathway (P -value = 0.0022), JAK-STAT signaling pathway (P -value = 0.0022) (Fig. 4g, h). Furthermore, the result of GSEA enrichment analysis showed that parent genes of DE ABS circRNAs are involved in some biological pathways, such as cell cycle, DNA repair, Ub-specific processing proteases, virus infection and Homologous recombination (Additional file 1: Fig S5c, d), suggesting that the DE ABS circRNAs play vital roles in CRC tumorigenesis.

The dysregulation network of splicing factors and DE ABS circRNAs

It has been reported that back-splicing is catalyzed by the canonical spliceosomal machinery and is modulated by *cis*-elements or *trans*-acting factors [32–34]. To investigate the potential regulation of ABS events by splicing factors (SFs) in CRC, we totally integrated previously published 67 SFs, which were all experimentally validated [19]. The result showed that a total of 54 SFs were significantly differentially expressed between CRC tumor and adjacent normal tissue samples (P -value < 0.05), among which 46 SFs were upregulated and 8 SFs were downregulated in CRC tumor samples (Additional file 2: Table S6). We calculated Spearman correlation coefficient between differentially expressed SFs and the PCU values of each DE ABS circRNA and filtered with the threshold of $|\rho| \geq 0.5$ and P -value < 0.05 . The high-confidence regulatory network contains 17 differentially expressed SFs (11 SFs

were upregulated and 6 SFs were downregulated, labeled circles) and 367 DE ABS circRNAs (Additional file 2: Table S7). This result indicated that DE ABS circRNAs could be regulated by SFs in CRC tumorigenesis.

In this regulatory network, a large fraction of DE ABS circRNAs ($n = 93$) were regulated by only one SF, whereas 32 DE ABS circRNAs were regulated by two different SFs (Fig. 5a). The number of regulated DE ABS circRNAs varied largely among different SFs, with the largest number for SRSF5 (154 DE ABS circRNAs) and the smallest number for QKI (1 DE ABS circRNA, Fig. 5b). Among 154 SRSF5 regulatory DE ABS circRNAs, 78 circRNAs have no other partners, and the rest circRNAs were also regulated by PCBP2 and KHSRP. Then, we examined the highly enriched pathways for each SF regulated sub-network (Additional file 1: Fig S6). For example, SRSF5 regulated DE ABS circRNAs were over-represented in Toll-like receptor signaling pathway, HIF-1 signaling pathway and TNF signaling pathway.

The prognostic value of DE ABS circRNAs in CRC

To evaluate the influence of clinical features on DE ABS circRNAs, we used the characteristics and histopathological features of CRC patients in our study, which were detailedly shown in Additional file 2: Table S8. We first detected the DE ABS circRNAs related to distant metastasis and 116 DE ABS circRNAs were found. Of which, 45 were A3BS circRNAs and 71 were A5BS circRNAs (Additional file 2: Table S9). Interestingly, all of these circRNAs were downregulated in patients with distant metastasis. We next grouped CRC patients into early-stage (stage I or II), late-stage (stage III or IV) groups and further explored the DE ABS circRNAs related to tumor stage. There were 12 DE A3BS and 25 DE A5BS circRNAs that were differentially expressed between early-stage and late-stage groups. Additionally, the PCU values of majority circRNAs (33 of 37) were decreased in late-stage group (Additional file 2: Table S10).

We next evaluated the overall survival rates of clinical features and those DE ABS circRNAs using the clinical information of CRC patients in our cohort. First, univariate Cox regression was performed to assess the relationship between clinicopathological features and the overall survival (OS) outcomes of patients in our CRC cohort. The results showed that distant metastasis (HR = 3.04, 95%CI 1.52–6.10; Wald test, P -value = 1.69×10^{-3}) and TNM stage (HR = 9.65, 95%CI 2.30–40.42; Wald test, P -value = 1.92×10^{-3}) were significantly associated with OS, while age and gender had no effect on patient's OS (Additional file 2: Table S11). The results of this assessment indicated that the survival data of our CRC cohort were informative and appropriate for further survival analysis. We next evaluated

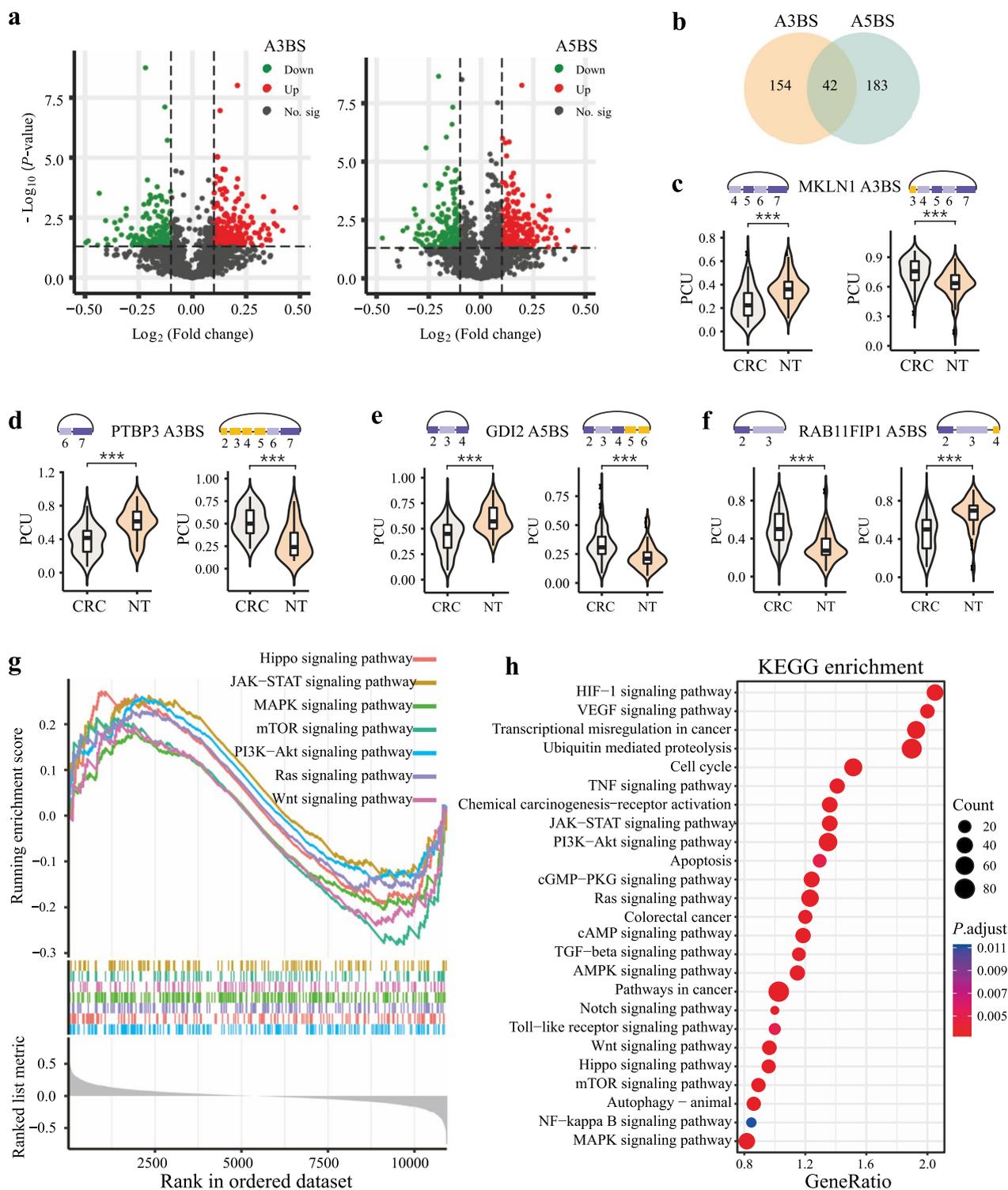


Fig. 4 Tumor-specific ABS events involved in CRC tumorigenesis. **a** Volcano plots showing differentially expressed A3BS and A5BS circRNAs between tumor and adjacent normal tissues. Green and red dots represent significantly down-regulated and up-regulated ABS circRNAs, respectively. **b** Intersection of parent genes of differentially expressed A3BS and A5BS circRNAs. **c-f** Violin plots showing the PCU values of A3BS and A5BS circRNAs in tumor and normal tissue samples. The top inset represents the genomic structure of ABS event. *** P -value < 0.001 . **g** GSEA analysis of parent genes of ABS circRNAs ranked by ΔPCU between CRC and normal tissues with Reactome pathway database. **h** KEGG enrichment analysis of parent genes of DE ABS circRNAs

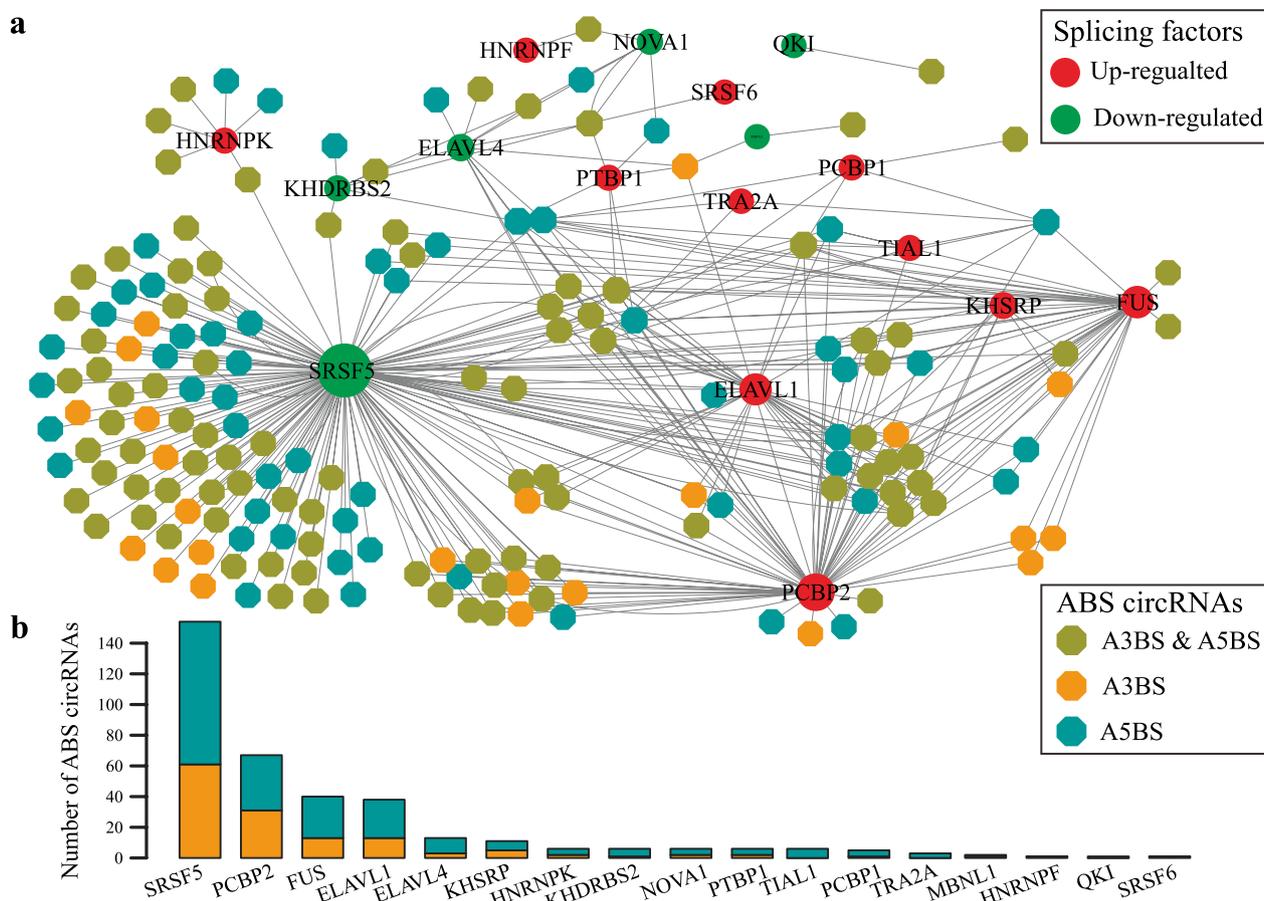


Fig. 5 The dysregulation network of splicing factors and ABS circRNAs. **a** The high-confidence regulation network of ABS events by splicing factors. Red circles represent up-regulated SF genes in CRC tumor samples, while green circles indicate down-regulated SF genes. Labeling size indicates the number of ABS circRNAs regulated by each SF. Colored octagon indicated ABS circRNAs regulated by SFs. **b** Number of ABS circRNAs regulated by each splicing factor. Color bars indicate the absolute number of each ABS type

the prognostic role of each DE ABS circRNA. As a result, a total of 32 ABS circRNAs from 27 genes were found to be significantly associated with patient overall survival (P -value < 0.05) in univariate Cox regression (Additional file 2: Table S12). Of these, 14 were A3BS circRNAs and 18 were A5BS circRNAs. All these prognostic related DE ABS circRNAs were unfavorable prognostic (Additional file 1: Fig S7). The most highly associated ABS circRNA with a decrease in overall survival is from *HIF1 A* gene, whose PCU value and RPM were increased in CRC tissues, suggesting that it might play a crucial role in CRC. Notably, all these 27 genes have been previously reported to play important roles in CRC, and 19 of these 27 genes are differentially expressed. Several genes are well-known components of cancer-related pathways, including mTOR signaling pathway (*HIF1 A*, *AKT3*), Toll-like receptor signaling pathway (*TRAF3*, *AKT3*), or Oxidative phosphorylation (*PPA2*) (Fig. 6a).

Subsequently, multivariate analysis was performed to further assess independent prognostic indicator in CRC using above 32 prognosis related DE ABS circRNAs. As shown in Fig. 6b, patient stage and 13 DE ABS circRNAs were both significant in univariate and multivariate analysis, which could be recognized as independent prognostic indicators for OS. Kaplan–Meier curves of stage and top 5 ABS circRNAs illustrated that patients stratifying according to the stage or PCU values had significantly different overall survival (Additional file 1: Fig S8). All above results indicated that DE ABS circRNAs could provide meaningful clinical values.

Different roles of *circXPO1-1* and *circXPO1-2* in CRC

Exportin 1 (*XPO1*, also known as chromosome region 1, *CRM1*) gene mediates the nuclear export of critical proteins, and has been reported to function as an oncogenic driver in cancer [35]. As shown in Fig. 3d, gene *XPO1* had an ABS event complexity up to 8. Moreover, we identified

all the circRNAs generated from *XPO1* gene were ABS circRNAs (Additional file 1: Fig S8a). Among these ABS circRNAs, *circXPO1-1* (239 nt) and *circXPO1-2* (371 nt) have relatively higher expression level (Additional file 1: Fig S8b). *circXPO1-1* back-spliced exon 3, exon 4 and exon 5 of *XPO1* gene with a length of 239 nucleotides (nt), and *circXPO1-2* back-spliced exon 2, exon 3, exon 4 and exon 5 of *XPO1* gene with a length of 371 nt (Additional file 1: Fig S8c). Subsequently, five cell lines were selected to investigate their existence. RT-PCR analysis revealed that *circXPO1-2* was more expressed in normal colon epithelial cells than in CRC cell lines. Sanger sequencing of PCR amplification products further validated the back-spliced junction sites (Fig. 7a). Moreover, divergent primers (to detect circRNAs) and convergent primers (to detect linear transcripts) were designed to validate the putative back-spliced junction site (Fig. 7b). These results are all confirmed the existence of these two ABS circRNAs in CRC cell lines. Compared with the linear *XPO1* mRNA transcript, *circXPO1-1* and *circXPO1-2* were resistant to RNase digestion, demonstrating the back-splicing nature of *circXPO1-1* and *circXPO1-2* (Fig. 7c). Since the function of circular RNA is always tightly associated with its subcellular location pattern [36], we analyzed the subcellular location of *circXPO1-1* and *circXPO1-2*. The results of qRT-PCR analysis after cell fractionation suggested that *circXPO1-2* is predominantly localized in the cytoplasm, whereas *circXPO1-1* is predominantly localized in the nucleus (Fig. 7d). Fluorescence in situ hybridization (FISH) assay with a probe spanning the back-spliced junction confirmed the subcellular localization of *circXPO1-1* and *circXPO1-2* (Fig. 7e). Therefore, we next examined the expression levels of *circXPO1-1* and *circXPO1-2* in 30-paired CRC samples. The qRT-PCR result showed that *circXPO1-2* was significantly down-regulated in CRC tumor tissues compared with matched adjacent non-tumor tissues, whereas *circXPO1-1* remained unchanged, which is consistent with the result derived from RNA-seq data (Fig. 7f). Additionally, the expression level of *circXPO1-2* was significantly lower in patients with TNM late-stage (Additional file 1: Fig S8 d). These evidences indicated that *circXPO1-2* is aberrantly expressed in CRC, and might play a role in CRC tumorigenesis and metastasis.

Functional assays were then conducted to further validate the role of *circXPO1-1* and *circXPO1-2* in CRC progression. Given that *circXPO1-2* was significantly down-regulated in CRC tissues, we individually transfected over-expression constructs into SW620 and LoVo cell lines. The result showed that the expression of circRNAs was specifically increased, and the expression of *XPO1* linear mRNA has no significant change (Additional file 1: Fig S9a). The results of CCK-8 and colony formation assay indicated that over-expression of *circXPO1-2* significantly suppresses cell proliferation ability of SW620 and LoVo cells (Fig. 8a, b). In parallel, *circXPO1-2* over-expression remarkably reduced cell migration and invasion abilities (Fig. 8c, d). Instead, overexpression of *circXPO1-1* had no significant effect on either cell proliferation or metastasis (Fig. 8a-d). The tumor volume and weight of over-expression of *circXPO1-2* were significantly lower than control and over-expression of *circXPO1-1* groups, which further confirmed the functional roles of *circXPO1-1/2* in vivo (Fig. 8e, f). This result suggested that *circXPO1-2* may play a tumor suppressor role in CRC, while *circXPO1-1* has no effect.

Next, we attempted to detect the underlying regulatory mechanism of *circXPO1-2*. *circXPO1-2* overexpression and control CRC cell lines were subjected to RNA-seq. A total of 314 up-regulated and 369 down-regulated genes were identified (Fig. 8g). Among them, we observed that several key positive regulators of AKT/ERK pathway were significantly down-regulated in *circXPO1-2* over-expression cells, including *VASP* (Vasodilator-stimulated phosphoprotein), and *KRT17* (Keratin 17). The result was further confirmed by qRT-PCR (Fig. 8h). KEGG enrichment analysis showed that those genes were over-represented in functional categories, such as cell cycle, MAPK signaling pathway, Hippo signaling pathway, Cellular senescence, NOD-like receptor signaling pathway *et. al.* (Fig. 8i). GSEA result showed that overexpression of *circXPO1-2* was under-represented in G2M checkpoint and over-represented in apoptosis and oxidative phosphorylation, which may explain the inhibition of CRC cell proliferation by *circXPO1-2* (Fig. 8j). Additionally, cancer hallmark pathways like G2M checkpoint, Oxidative Phosphorylation, E2 F Targets, Interferon Alpha

(See figure on next page.)

Fig. 6 ABS circRNAs associated with unfavorable prognosis in CRC. **a** ABS circRNAs significantly associated with patient overall survival by univariate analysis. ABS circRNAs with $|\Delta\text{PCU}| > 0.1$ and with significant survival association are shown, ranked by differential survival. ABS circRNAs are labelled with gene name, number of detected patients, ABS event type and prognostic type. Other related information for each ABS circRNA is exhibited in heatmaps, including survival prognosis, tissues PCU values, and RPM expression levels. **b** ABS circRNAs significantly associated with patient overall survival by univariate analysis and multivariate analysis

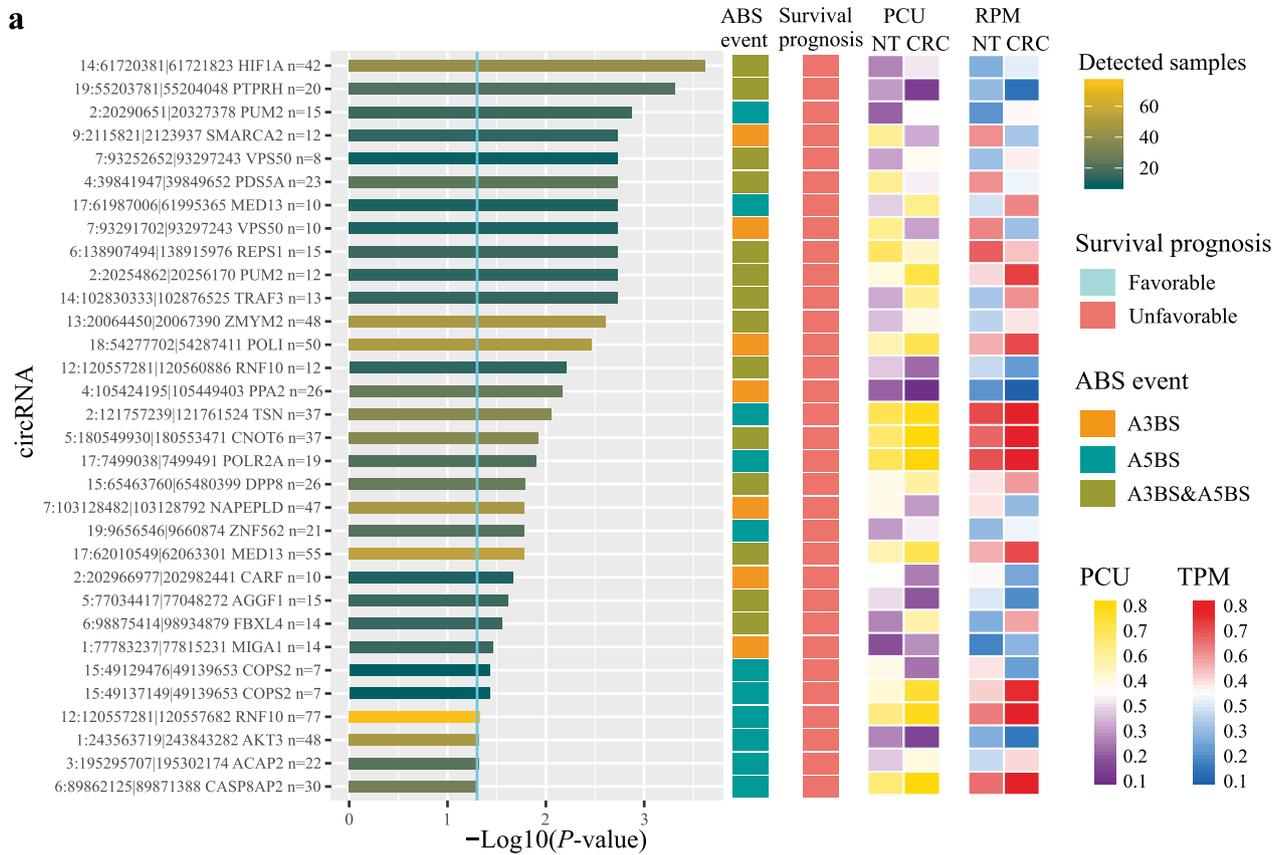


Fig. 6 (See legend on previous page.)

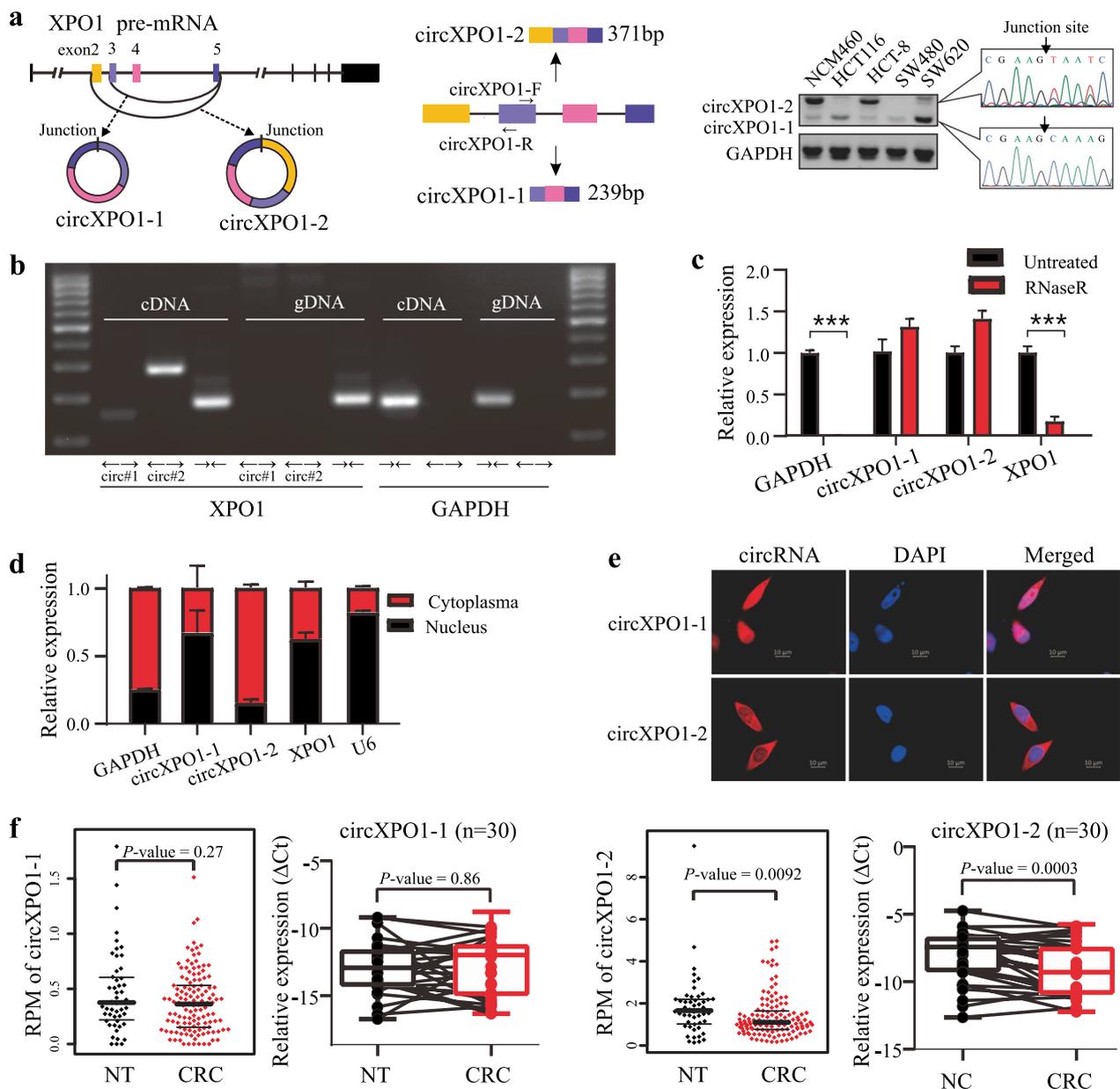


Fig. 7 Identification of *circXPO1-1* and *circXPO1-2* in CRC. **a** Schematic illustration and validation of an A3BS event of *XPO1* gene. The A3BS event contains two ABS circRNAs, *circXPO1-1* and *circXPO1-2*. The existence of *circXPO1-1* and *circXPO1-2* was proved by RT-PCR using divergent primer in common region in CRC cell lines and their back-splicing junction sites were verified by Sanger sequencing. **b** Agarose gel electrophoresis analysis of qPCR products of divergent primer and convergent primer. ← → divergent primers; → ← convergent primers. cDNA: complementary DNA, gDNA: genomic DNA. GAPDH was used as the negative control. **c** Relative expression level of *circXPO1-1*, *circXPO1-2* and *XPO1* in SW620 cell line after RNase R treatment. *** *P*-value < 0.001. **d** qPCR result showing the different subcellular location of *circXPO1-1* and *circXPO1-2*. *circXPO1-2* is predominantly localized in the cytoplasm, whereas *circXPO1-1* is predominantly localized in the nucleus. GAPDH and U6 were used as the positive control of cytoplasm location and nucleus location, respectively. **e** FISH assay showing the subcellular location of *circXPO1-1* and *circXPO1-2*. **f** *circXPO1-1* and *circXPO1-2* expression measured by RNA-seq data and by qRT-PCR method in additional CRC cohort, respectively

Response, P53 Pathway were modulated by *circXPO1-2* over-expression (Additional file 1: Fig S9b).

Discussion

CircRNAs have been proved to be a class of endogenous RNAs with covalently closed-loop structures generated by back-splicing [1, 2]. With the development of

next-generation sequencing technology, circRNAs are widely detected in human. circRNAs have been reported to be essential regulators in various biological processes through diverse molecular mechanisms, such as sponging microRNA (miRNA), acting as RNA-binding protein (RBP) decoy, and translating into proteins [6, 10, 16]. Mounting evidence has been revealed that circRNAs can be aberrantly expressed in human cancer and play important roles during tumorigenesis or metastasis [30, 32, 33, 37, 38]. One important molecular mechanism for circRNAs is alternative back-splicing, through which multiple circRNAs sharing the same 3' back-splice site or 5' back-splice site can be produced from a gene locus. The alternative back-splicing largely expands the complexity of transcriptome. Pioneer works have reported ABS events in human cell lines, and they are still rarely reported in human cancer. In this work, we analyzed rRNA-depleted transcriptome sequencing data in human CRC samples, and comprehensively characterized their landscape of ABS events. Some aberrantly expressed ABS circRNAs have been detected between CRC and adjacent normal tissues, shedding light on the importance of these ABS circRNAs.

It has been documented that the biogenesis of circRNAs through back-splicing can be regulated by both *cis*- or *trans*-acting factors [6]. Consistent with previous work [16], we also found that the competition between *Alu* element pairs across flanking introns promotes the ABS occurrence. In addition, our result showed that longer flanking introns were correlated with higher complexity of ABS events. These findings reinforce the evidence that *cis*-regulatory elements contribute greatly to ABS events. In addition, differential expression of splicing factors can trigger the differences of ABS patterns. Integrating the expression of splicing factors and differentially expressed ABS circRNAs can provide an approach to address the underlying molecular mechanisms of ABS pathways involved in CRC. We built a dysregulation network of splicing factors and differentially expressed ABS circRNAs. In this regulatory network, some splicing factors, such as *FUS*, *ELAVL4*, *KHSRP*, *HNRNP*, *MBNL1* and *QKI* have been reported to have the ability to recognize specific motifs and participate in the biogenesis of

circRNAs [32, 37, 39–42], and *PCBP2* and *ELAVL1* can interact with circRNAs to be involved in tumorigenesis [39, 43]. These results further proved the importance of the *cis*- or *trans*-regulatory mechanism in CRC.

Previous studies had revealed that the more an exon is circularized, the less it is represented in mature mRNA [44], which indicated that there may exist linear-circular switching [3]. Although the biological functions of circRNAs are not fully uncovered, the competitive splicing between circRNA and their linear counterparts would help us understand the mechanism of circRNAs involved in carcinogenesis. Therefore, systematic characterization of differentially expressed ABS circRNAs in CRC will help us understand how circRNAs might contribute to tumorigenesis. In this work, a total of 552 ABS circRNAs were identified to be differentially expressed between CRC and adjacent normal tissues. Some cancer-related functional categories, such as mTOR signaling pathway, p53 signaling pathway, PI3 K-Akt signaling pathway, Ras signaling pathway and Wnt/ β -catenin signaling pathway, have been identified to be highly enriched in these genes that occurred ABS circRNAs. This result suggested that competitive splicing with their linear counterparts is an important mechanism for circRNAs to contribute to disease progression.

Elucidation of the functional implications and clinical relevance of such dysregulated ABS circRNAs is helpful to identify new diagnostic biomarkers and therapeutic targets. Univariate and multivariate analysis illustrated that 13 DE ABS circRNAs could act as independent prognostic indicators for CRC patients. Interestingly, we found two novel ABS circRNAs, *circXPO1-1* and *circXPO1-2*, which have different functional roles in CRC. *circXPO1-2* is down-regulated and serves as a tumor suppressor in CRC, while the expression of *circXPO1-1* has no significant change. Further research is needed to completely elucidate their transcriptional regulation and processing of transcriptional output of this locus. Nevertheless, XPO1 inhibitor, Selinexor, has been approved for the treatment of relapsed multiple myeloma. Considering the tumor suppressor role of *circXPO1-2* and the oncogene role of *XPO1*, it is a new strategy to treat CRC

(See figure on next page.)

Fig. 8 Different functional roles of *circXPO1-1* and *circXPO1-2* in HCC cell lines. **a** CCK-8 assay and **(b)** Colony formation assay in SW620 and LoVo cells detecting cell proliferation transfected with control, OE-*circXPO1-1*, and OE-*circXPO1-2* vectors. **c, d** Transwell assays reveal cell migration and invasion in SW620 and LoVo cells transfected with control, OE-*circXPO1-1*, and OE-*circXPO1-2* vectors. **e** Tumor size and tumor weight **(f)** were accessed for each derived xenograft tumor. Xenograft tumor models were established using LoVo cells stably transfected with control, OE-*circXPO1-1*, and OE-*circXPO1-2* vectors. **g** Volcano plot of differentially expressed genes in LoVo cells stably transfected with control and OE-*circXPO1-2* vectors. **h** Relative mRNA expression levels of differentially expressed genes detected by qRT-PCR in control and OE-*circXPO1-2* LoVo cells. ** *P*-value < 0.01, *** *P*-value < 0.001. **i** Cancer hallmark enrichment analysis of differentially expressed genes in control and OE-*circXPO1-2* LoVo cells. **j** GSEA plots of differentially expressed gene sets involved in regulation of carcinogenesis

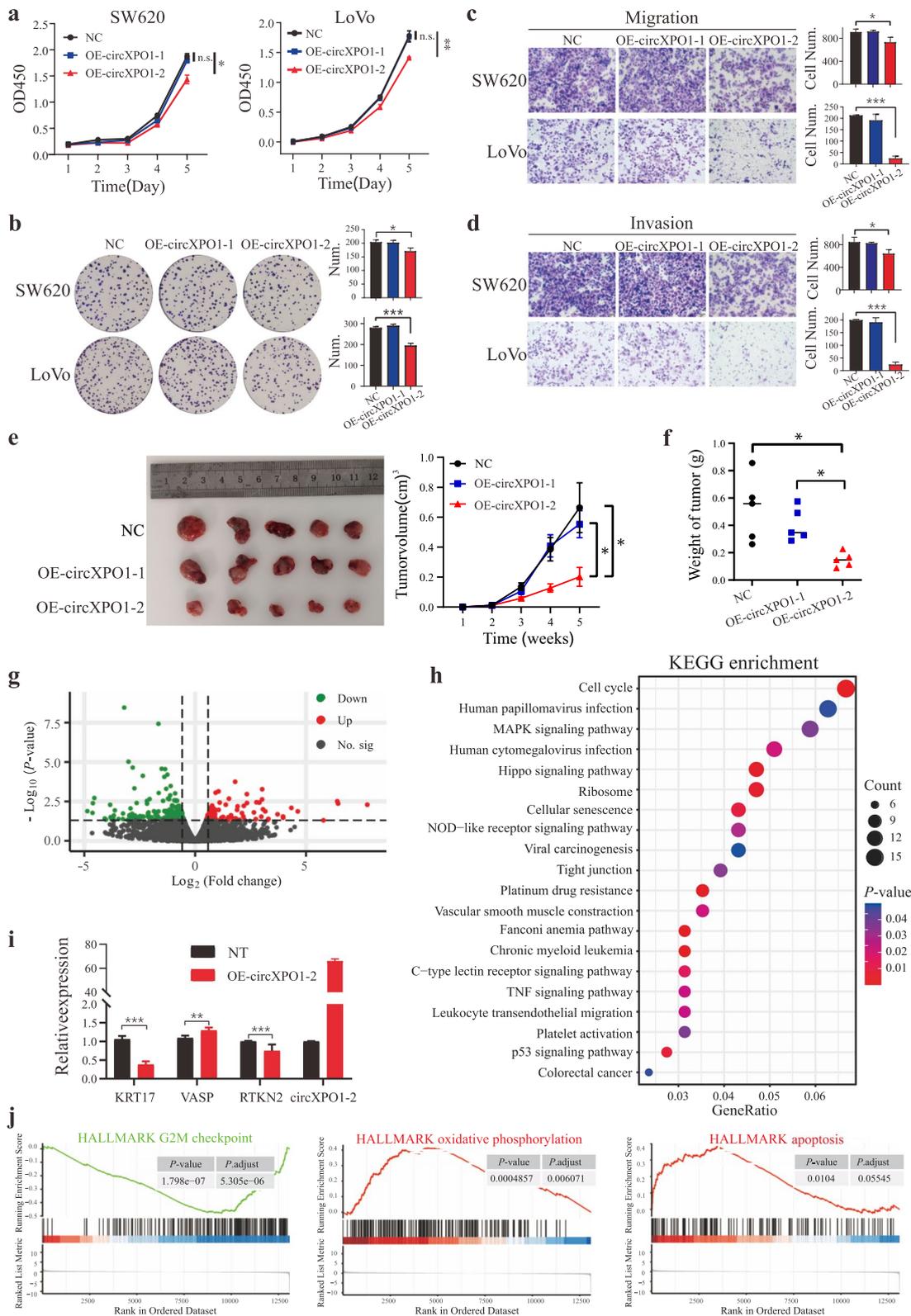


Fig. 8 (See legend on previous page.)

patients inducing splicing-switch from linear isoform to circulation isoform.

Through Sanger sequencing, we found *circXPO1-1* and *circXPO1-2* are generated from a newly annotated transcript of *XPO1*, ENST00000680228, but not the canonical transcript encoding the functional protein. Because of a 64 bp length insertion, this transcript carries a premature stop codon, likely undergoing nonsense mediated decay. This further supported the competitive splicing roles of circRNAs and their linear counterparts. Much more, it also indicated that detection of junction site is only the first step of circRNA identification. How many exons and which exons are included in mature circRNAs are further to be validated. Fully considering the alternative splicing and alternative back-splicing of circRNAs and their linear counterparts is needed when exploring their functional roles.

Conclusions

In conclusion, we comprehensively described the landscape of ABS events and concluded the preference of ABS events in CRC. Our results demonstrated that differentially expressed ABS circRNAs were highly related to cancer hallmarks and highlighted their important roles in cancer biology. Our work offers a new set of targets with potential clinical benefits.

Abbreviations

| | |
|----------|----------------------------------|
| circRNAs | Circular RNAs |
| ABS | Alternative back-splicing events |
| CRC | Colorectal cancer |
| A3BS | Alternative 3 back-splicing |
| A5BS | Alternative 5 back-splicing |
| RPM | Reads per million mapped reads |
| PCU | Percent circularized-site usage |
| SF | Splicing factors |
| GSEA | Gene set enrichment analysis |
| TS | Tumor suppressor |
| DE ABS | Differentially expressed ABS |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06479-2>.

Additional file 1

Additional file 2

Acknowledgements

We thank Professor Chunqing Guo for insightful discussions.

Author contributions

Q.N., and D.D. conceived and designed the study; Q.N., and L.Z. performed data collection and analysis of raw data; Q.N., Q.J., Y.W., J.W., and L.Y. conducted the experimental validation; Q.J. performed the animal experiments; Q.N. analyzed data, interpreted results, and prepared figures; Y.H., and Q.Z. performed sample preparation; Q.N., and D.D. drafted the manuscript; Q.N., and F.C. revised the manuscript; all authors reviewed and edited the final version of the manuscript.

Funding

This project is supported by grants from the National Natural Science Foundation of China (32470593, 82072819), and the Natural Science Research Project in Higher Education of Jiangsu Province (22 KJB320025).

Availability of data and materials

The RNA-seq data of 88 CRC tissue samples and overexpression *circXPO1-2* in LoVo cell have been deposited into the NCBI Sequencing Read Archive database under the accession numbers PRJNA694010 and PRJNA945510, respectively.

Declarations

Ethics approval and consent to participate

A total of 78 CRC patients were participated in this research with informed consent and this study was conducted following approval by the ethical committee of the First Affiliated Hospital of Soochow University.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Cancer Institute, Xuzhou Medical University, 209 Tongshan Road, Xuzhou 221004, Jiangsu, China. ²Center of Clinical Oncology, The Affiliated Hospital of Xuzhou Medical University, 99 Huaihai Road, Xuzhou 221002, Jiangsu, China. ³Jiangsu Center for the Collaboration and Innovation of Cancer Biotherapy, Xuzhou Medical University, 209 Tongshan Road, Xuzhou 221004, Jiangsu, China. ⁴Shanghai YangZhi Rehabilitation Hospital (Shanghai Sunshine Rehabilitation Center), School of Medicine, Tongji University, Shanghai, China. ⁵Department of General Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China.

Received: 7 December 2023 Accepted: 11 April 2025

Published online: 24 April 2025

References

- Liu CX, Chen LL. Circular RNAs: characterization, cellular roles, and applications. *Cell*. 2022;185:2016–34.
- Kristensen LS, Jakobsen T, Hager H, Kjems J. The emerging roles of circRNAs in cancer and oncology. *Nat Rev Clin Oncol*. 2022;19:188–206.
- Zhang J, Chen S, Yang J, Zhao F. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun*. 2020;11:90.
- Zhang J, Hou L, Zuo Z, Ji P, Zhang X, Xue Y, Zhao F. Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat Biotechnol*. 2021;39:836–45.
- Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol*. 2014;32:453–61.
- Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, Chen LL, Yang L. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res*. 2016;26:1277–87.
- Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genome-wide characterization of non-polyadenylated RNAs. *Genome Biol*. 2011;12:R16.
- Chuang TJ, Chen YJ, Chen CY, Mai TL, Wang YD, Yeh CS, Yang MY, Hsiao YT, Chang TH, Kuo TC, et al. Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-backsplicing in human cells. *Nucleic Acids Res*. 2018;46:3671–91.
- Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*. 2013;19:141–57.
- Lei M, Zheng G, Ning Q, Zheng J, Dong D. Translation and functional roles of circular RNAs in human cancer. *Mol Cancer*. 2020;19:30.

11. Chen S, Shen X. Long noncoding RNAs: functions and mechanisms in colon cancer. *Mol Cancer*. 2020;19:167.
12. Yang H, Li X, Meng Q, Sun H, Wu S, Hu W, Liu G, Li X, Yang Y, Chen R. CircPTK2 (hsa_circ_0005273) as a novel therapeutic target for metastatic colorectal cancer. *Mol Cancer*. 2020;19:13.
13. Wang X, Zhang H, Yang H, Bai M, Ning T, Deng T, Liu R, Fan Q, Zhu K, Li J, et al. Exosome-delivered circRNA promotes glycolysis to induce chemoresistance through the miR-122-PKM2 axis in colorectal cancer. *Mol Oncol*. 2020;14:539–55.
14. Zheng X, Chen L, Zhou Y, Wang Q, Zheng Z, Xu B, Wu C, Zhou Q, Hu W, Wu C, Jiang J. A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling. *Mol Cancer*. 2019;18:47.
15. Zheng G, Ma Y, Zou Y, Yin A, Li W, Dong D. HCMDb: the human cancer metastasis database. *Nucleic Acids Res*. 2018;46:D950–5.
16. Zhang P, Zhang XO, Jiang T, Cai L, Huang X, Liu Q, Li D, Lu A, Liu Y, Xue W, et al. Comprehensive identification of alternative back-splicing in human tissue transcriptomes. *Nucleic Acids Res*. 2020;48:1779–89.
17. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90.
18. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–15.
19. Xiong Y, Deng Y, Wang K, Zhou H, Zheng X, Si L, Fu Z. Profiles of alternative splicing in colorectal cancer and their clinical significance: a study based on large-scale sequencing data. *EBioMedicine*. 2018;36:183–95.
20. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:147–25.
21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7.
22. Dressler L, Bortolomeazzi M, Keddar MR, Missetic H, Sartini G, Achasagredo A, Montorsi L, Wijewardhane N, Repana D, Nulsen J, et al. Comparative assessment of genes driving cancer and somatic evolution in non-cancer tissues: an update of the Network of Cancer Genes (NCG) resource. *Genome Biol*. 2022;23:35.
23. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT StringTie and Ballgown. *Nat Protoc*. 2016;11:1650–67.
24. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multi-factor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40:4288–97.
25. Yang Y, Zhao Y, Zhang W, Bai Y. Whole transcriptome sequencing identifies crucial genes associated with colon cancer and elucidation of their possible mechanisms of action. *Onco Targets Ther*. 2019;12:2737–47.
26. Wu M, Li W, Huang F, Sun J, Li KP, Shi J, Yang J, Li J, Li Y, Hu N, Hu Y. Comprehensive analysis of the expression profiles of long non-coding RNAs with associated ceRNA network involved in the colon cancer staging and progression. *Sci Rep*. 2019;9:16910.
27. Ju HQ, Zhao Q, Wang F, Lan P, Wang Z, Zuo ZX, Wu QN, Fan XJ, Mo HY, Chen L, et al. A circRNA signature predicts postoperative recurrence in stage II/III colon cancer. *EMBO Mol Med*. 2019;11: e10168.
28. Meng X, Hu D, Zhang P, Chen Q, Chen M. CircFunBase: a database for functional circular RNAs. *Database (Oxford)*. 2019;2019:003.
29. Zhang XT, Liu TL, Huang JL, He JP. PICALM exerts a role in promoting CRC progression through ERK/MAPK signaling pathway. *Cancer Cell Int*. 2022;22(1):178.
30. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*. 2014;159:134–47.
31. Dong R, Ma XK, Chen LL, Yang L. Increased complexity of circRNA expression during species evolution. *RNA Biol*. 2017;14:1064–74.
32. Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell*. 2014;56:55–66.
33. Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung LH, Bindereif A. Exon circularization requires canonical splice signals. *Cell Rep*. 2015;10:103–11.
34. Chen LL. The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol*. 2016;17:205–11.
35. Azizian NG, Li Y. XPO1-dependent nuclear export as a target for cancer therapy. *J Hematol Oncol*. 2020;13:61.
36. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*. 2016;17:47–62.
37. Conn SJ, Pillman KA, Toubia J, Conn VM, Salamanidis M, Phillips CA, Roslan S, Schreiber AW, Gregory PA, Goodall GJ. The RNA binding protein quaking regulates formation of circRNAs. *Cell*. 2015;160:1125–34.
38. Zhao W, Cui Y, Liu L, Qi X, Liu J, Ma S, Hu X, Zhang Z, Wang Y, Li H, et al. Splicing factor derived circular RNA circUHRF1 accelerates oral squamous cell carcinoma tumorigenesis via feedback loop. *Cell Death Differ*. 2020;27:919–33.
39. Chen J, Wu Y, Luo X, Jin D, Zhou W, Ju Z, Wang D, Meng Q, Wang H, Fu X, et al. Circular RNA circRHOBTB3 represses metastasis by regulating the HuR-mediated mRNA stability of PTBP1 in colorectal cancer. *Theranostics*. 2021;11:7507–26.
40. Dell'Orco M, Elyaderani A, Vannan A, Sekar S, Powell G, Liang WS, Neisewander JL, Perrone-Bizzozero NI. HuD regulates mRNA-circRNA-miRNA networks in the mouse striatum linked to neuronal development and drug addiction. *Biology (Basel)*. 2021;10(9):939.
41. Okholm TLH, Sathe S, Park SS, Kamstrup AB, Rasmussen AM, Shankar A, Chua ZM, Fristrup N, Nielsen MM, Vang S, et al. Transcriptome-wide profiles of circular RNA and RNA-binding protein interactions reveal effects on circular RNA biogenesis and cancer pathway expression. *Genome Med*. 2020;12:112.
42. Razpotnik R, Nassib P, Kunej T, Rozman D, Rezen T. Identification of novel RNA binding proteins influencing circular RNA expression in hepatocellular carcinoma. *Int J Mol Sci*. 2021;22(14):477.
43. Chen Y, Ling Z, Cai X, Xu Y, Lv Z, Man D, Ge J, Yu C, Zhang D, Zhang Y, et al. Activation of YAP1 by N6-methyladenosine-modified circCPSF6 drives malignancy in hepatocellular carcinoma. *Cancer Res*. 2022;82:599–614.
44. Kelly S, Greenman C, Cook PR, Papantonis A. Exon skipping is correlated with exon circularization. *J Mol Biol*. 2015;427:2414–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.