Human Genomics

RESEARCH

Open Access

Exon skipping as a potential diagnostic biomarker in colorectal cancer: an integrated epigenomic-transcriptomic analysis

Lili Zhang[1†], Jian Cui[2†], Jinxin Shi[2†], Jiahui Cai[1], Tianhan Sun[2], Gaoyuan Sun[1], Yifei Li[1], Hexin Li[1], Siyuan Xu[1], Xiaokun Tang[1], Ziwei Chen[3], Hongyuan Cui[2*], Fei Xiao[1*] and Gang Zhao[2*]

Abstract

Background Colorectal cancer (CRC) is the third most common cancer globally. Alternative splicing contributes significantly to CRC tumorigenesis through aberrant transcript generation. However, the regulatory influence of RNA modifications on splicing remains poorly understood, largely due to technical difficulty. Nanopore direct RNA sequencing addresses this by enabling simultaneous detection of RNA modifications and Alternative splicing events (ASEs).

Methods We conducted Nanopore direct RNA sequencing on paired tumor and normal tissues from surgical resections at Beijing Hospital. Differential putative RNA modification sites and ASEs linked to CRC were systematically identified. To validate the key findings, we utilized a large patient cohort from The Cancer Genome Atlas (TCGA) and predicted 3D protein structures with AlphaFold3. The predicted structures were then compared using TM-align. Regulatory relationships between RNA modifications and splicing were explored through predictive modeling of potential cis-regulatory pairs. The splicing events were also validated.

Results The MYH11-201 transcript of the MYH11 gene contains an additional exon (ENSE00001632812) compared to the MYH11-203 isoform. Both bioinformatic analysis and experimental validation confirmed frequent loss of this exon in tumor tissues. This finding was further validated in the TCGA cohort, demonstrating a significant preference for exon skipping in tumor tissues. These results suggest that the skipping of ENSE00001632812 is a promising candidate biomarker associated with CRC pathogenesis. Notably, this exon's PF00063 domain interacts with multiple tumor suppressor genes and oncogenes domains, suggesting its functional importance. The structures revealed pronounced rotational divergence within a putative C-terminal transmembrane domain-like region. Furthermore, we utilized Nanopore sequencing to explore the potential interplay between alternative splicing and RNA modifications.

†Lili Zhang, Jian Cui and Jinxin Shi have contributed equally to this work.

*Correspondence:
Hongyuan Cui
cuihongyuan3921@bjhmoh.cn
Fei Xiao
xiaofei3965@bjhmoh.cn
Gang Zhao
zhaogang7563@163.com

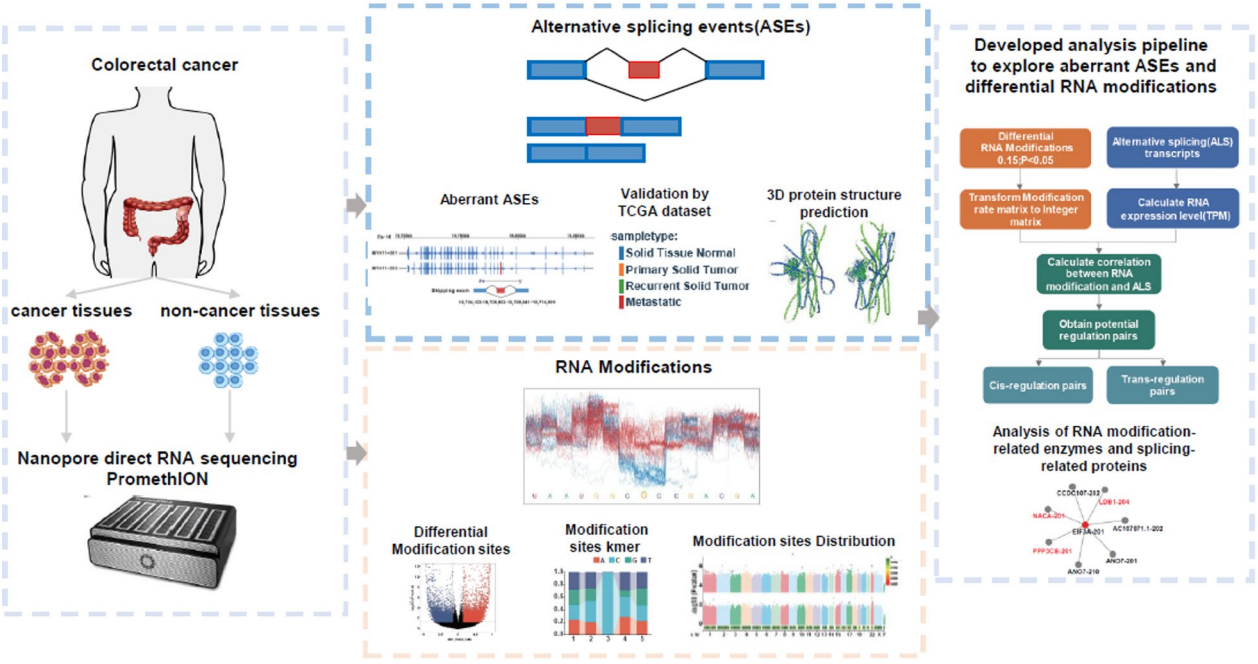
Full list of author information is available at the end of the article

We implemented an integrated analytical workflow (available at https://github.com/lelelililele/Nanopore-ASEs-and-RNA-modification) combining modification calling and splicing analysis tools to investigate RNA modification-related enzymes and splicing-related proteins in CRC.

Conclusions This pilot study utilizes Nanopore direct RNA sequencing to characterize exon skipping events and RNA modifications in CRC. We identified the skipping of MYH11 exon ENSE00001632812 as a potential candidate for future diagnostic investigation. By integrating modification and splicing data, we highlighted putative regulatory pairs that warrant further functional exploration. While our findings offer new insights into CRC molecular mechanisms, extensive validation in independent large-scale cohorts and functional assays is essential to confirm the diagnostic utility and mechanistic roles of these targets.

Keywords Exon skipping, RNA modifications, Alternative splicing events, Colorectal cancer, Nanopore sequencing

Graphical Abstract



Introduction

Alternative splicing of precursor messenger RNA (pre-mRNA) enhances the complexity of gene regulation and transcript diversity, thereby influencing both physiological and pathological processes [1]. Variable cis-regulatory elements and alterations in splicing trans-regulatory factors modulate ASEs, which aberrant ASEs impact CRC progression by regulating cell proliferation, invasion, apoptosis, as well as angiogenesis and drug resistance [2, 3]. For instance, serine and arginine-rich splicing factor 1 (SRSF1), a key member of the SR family, has been observed to be upregulated in CRC and is correlated with DNA damage and cell cycle progression [4]. SRSF1 acts as an oncogenic factor that promotes tumorigenesis by regulating various ASEs, such as the splicing of Rac Family Small GTPase 1 (RAC1) [2]. Although alternative splicing plays a crucial role in CRC progression and is implicated in multiple malignant hallmarks, its function remains less understood compared to its involvement in prostate or breast cancer. The specific splicing alterations associated with disease are not yet fully characterized, underscoring the need for further research to comprehensively elucidate the regulatory mechanisms governing ASEs in CRC [5, 6].

Over 150 types of RNA modifications have been confirmed to regulate gene expression, extending beyond ribosomal and transfer RNAs to mRNAs and non-coding RNAs [7–9]. Dysregulation of the epigenetic machinery—comprising 'writers,' 'erasers,' and 'readers'—is implicated in various diseases, particularly cancer [10–13]. While traditional detection methods face limitations in sensitivity and quantification [20], Oxford Nanopore direct RNA sequencing overcomes these challenges by enabling direct, transcriptome-wide profiling of modifications through electrical signal analysis [14].

Previous studies demonstrated the regulatory role of RNA modifications in post-transcriptional events [15]. For instance, the absence of the m6A methyltransferase METTL3 significantly alters the alternative splicing patterns of thousands of genes [16–20]. During alternative splicing, m6A is enriched in both exons and introns, with m6A-modified exons showing a tendency to be retained in mRNA following splicing [16, 18, 19]. These findings underscore the critical influence of RNA modifications on alternative splicing. Conversely, ASEs can also modulate m6A modifications. One study identified METTL3-D, an alternatively spliced isoform of the METTL3 protein that lacks expression. This isoform expressed with higher levels in normal liver tissues compared to the canonical full-length METTL3-A, but it is significantly reduced in tumor tissues and cell lines [21]. Functional experiments further demonstrated that METTL3-D expression leads to decreased m6A modifications, which inhibit the proliferation, migration, and invasion of hepatocellular carcinoma cells, and negatively correlates with tumor malignancy in patients, exhibiting effects opposite to those of METTL3-A [21].

While the regulatory role of RNA modifications in splicing is recognized, accurately mapping these interactions on native RNA molecules remains challenging. Therefore, our research employed Nanopore direct RNA sequencing to simultaneously profile differential RNA modification sites and aberrant ASEs in CRC, providing an integrated perspective on their potential regulatory relationships.

Result

Aberrant ASEs in CRC

Compared to short read sequencing, Nanopore sequencing significantly improved accuracy in detecting aberrant ASEs. For paired CRC tissues, the relative abundances of splicing events were quantified using the proportion spliced-in (PSI), with ΔPSI representing the differences in these relative abundances between conditions (Fig. 1A). A total of 263 aberrant ASEs (Table S1) were identified during the screening process, as indicated in Fig. 1B. These events were predominantly associated with exon skipping and alternative first exons, accounting for 37.26% and 23.95% respectively (Fig. 1C). Among the identified ASE genes, notable genes linked to Wnt signaling pathway included *PPP3CB* [22], *CAMK2G* [23], *GSK3B* [24], *RAC1* [25], and *CTNNBIP1* [26]. Additionally, CRC related genes such as *CORO1C* [27], *COL1A2* [28], *FN1* [29], *TPM1* [30], and *MYH11* [31] were also detected in this analysis (Fig. 1C). These ASE genes were found to be enriched in pathways related to key cellular processes including migration, proliferation, differentiation, angiogenesis, and membrane functions (Fig. 1D). Furthermore, a network combining protein-protein interactions and protein domain-domain interactions was constructed (Fig. 1E), with *RPL18A* and *RPS24* identified as the top nodes in terms of connectivity. Both of these genes are associated with ribosomal functions, and the alternative splicing of *RPS24* has been implicated as a marker of cancer progression and epithelial-mesenchymal transition [32]. Other cancer-related genes such as *GSK3B* [24], *FN1* [29], *MYH11* [31], *TPM1* [30], and *NQO1* [33] were observed to exert significant influence within the network.

Aberrant ASEs gene *MYH11* in CRC

For ASE genes, we directed our focus towards the CRC-associated gene *MYH11*, which has shown a correlation with patients in stage II and III CRC [31]. Out of the 13 transcripts of *MYH11*, the primary protein-coding transcripts were MYH11-201 and MYH11-203(Fig. 2A and B). A decrease in the expression of the *MYH11* gene was observed in tumor tissues (Fig. 2C), along with a corresponding decrease in the TPM levels of MYH11-201 and MYH11-203(Fig. 2D). To investigate the structural consequences of exon variation, we predicted the 3D structures of both isoforms with AlphaFold3 and compared them using TM-align. Comparative analysis revealed a pronounced rotational divergence within a putative C-terminal transmembrane domain-like region (Fig. 2E). This predicted structural alteration provides a hypothesis for potential functional divergence between the two isoforms. Interestingly, despite this overall decrease, the proportion of MYH11-201 transcripts exhibited a significant increase in the tumor samples, while the levels of MYH11-203 remained decreased (Fig. 2B, F, G).

Distinct from MYH11-201, MYH11-203 features a unique exon, ENSE00001632812, which is linked to the PF00063 domain (Figs. 2A and 3A). Through ASE screening, we uncovered an aberrant splicing event occurring in this exon of MYH11-203 (Fig. 3A), resulting in a tendency to exon skipping in tumor tissues. Moreover, there was a notable reduction in the proportion of ENSE00001632812 in the tumor (Fig. 3B). At the gene level, *MYH11* expression appeared to be downregulated in tumor samples. Despite this downregulation, at the transcript level, the proportion of MYH11-201 showed an increase within tumor tissue, while the proportion of MYH11-203 decreased. Additionally, at the exon level, the proportion of ENSE00001632812 within MYH11-203 was observed to decrease in tumor samples. This prompts further investigation into the functional implications of the skipped exon in *MYH11*. We also validated changes in *MYH11* splicing by RT-PCR (Fig. 3C). Due to our limited sample size, we further validated this finding using TCGA COAD dataset, which revealed that the usage and expression of this exon was significantly higher in normal tissues compared to tumor tissues (Fig. 3D and

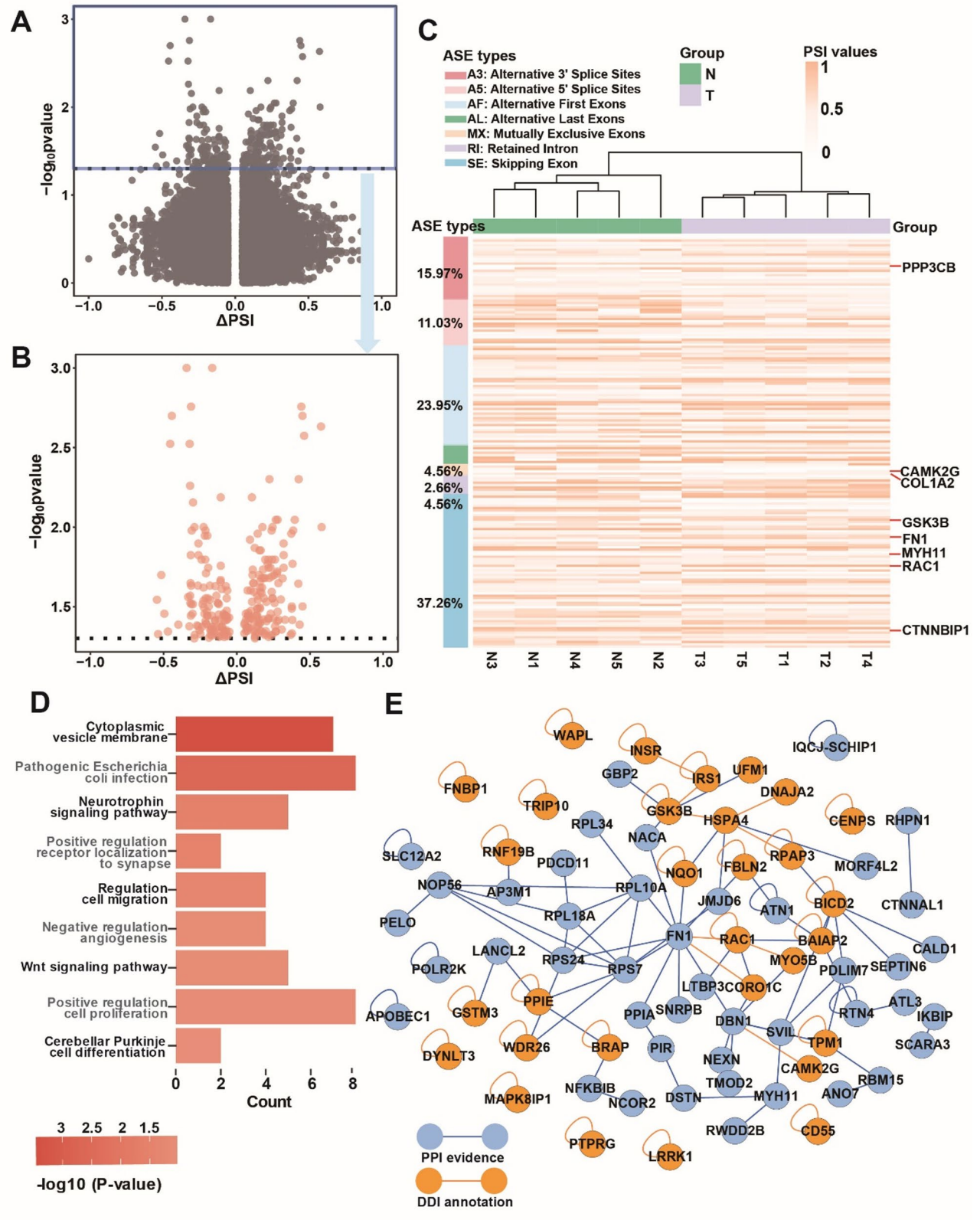


Fig. 1 Aberrant ASEs in CRC **A**. Volcano plot of ΔPSI between CRC paired tissues. Each point represents an alternative splicing event **B**. Only P-value < 0.05 genes was considered **C**. Heatmap of Genes with aberrant ASEs. The PSI values were used to classified samples. The ASE types classification were exhibited in left part **D**. Functional enrichment analysis of genes with aberrant alternative splicing **E** PPI and DDI network between aberrant alternative splicing genes. PSI: proportion spliced-in. ASE: alternative splicing event. ΔPSI: between Cond1 and Cond2, ΔPSI = PSI_2 - PSI_1. PPI: protein-protein interactions DDI: domain-domain interactions

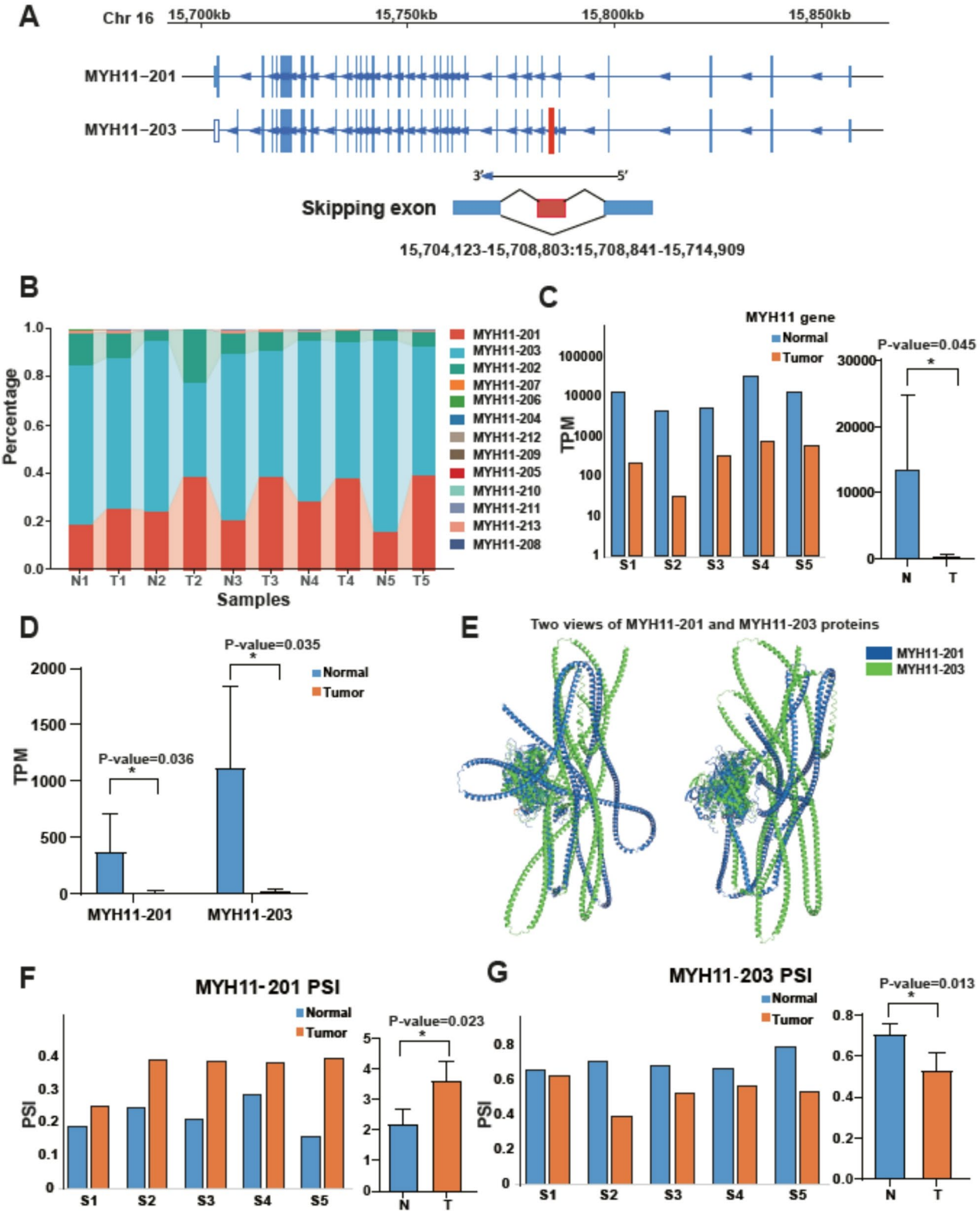


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Aberrant ASEs gene MYH11 in CRC **A**. Structural representation of MYH11-201 and MYH11-203 isoforms, with the red exon indicating the aberrantly skipped exon in CRC, located on the reverse strand of chr16 spanning coordinates 15,708,803 to 15,708,841 **B**. Distribution of all MYH11 transcript isoforms, totaling 13 transcripts **C**. Expression levels of the MYH11 gene in paired tissues, showing higher expression in normal tissues compared to tumor tissues. The blue bars for normal tissues, orange bars for tumor tissues. S1-S5:sample1-sample5 **D**. TPM values of MYH11-201 and MYH11-203 in paired samples, both showing lower expression in tumor tissues. TPM: Transcripts Per Million **E** 3D protein structures of MYH11-201 and MYH11-203 isoforms. It reveals conformational divergence in the coiled-coil tail **F**. **G**. PSI values of MYH11-201 and MYH11-203, with MYH11-203 exhibiting lower relative expression in tumor tissues compared to normal tissues, while MYH11-201 shows the opposite trend

E). It prompted us to further investigate the functional domain of this exon.

Exploring domain-domain interactions (Fig. 3F), we identified that the exon related PF00063 domain exerts influence over multiple domains in cancer-associated genes, including tumor suppressor genes such as *INF2* [34] and *POC2A* [35], as well as tumor driver genes like *MYOSA* [36], *MYO1E* [37], *ACTR3* [38], and *ACTB* [39]. Notably, the dual-function gene *DYRK1A* [40], acting as both a tumor suppressor and an oncogene, was also impacted by this domain interaction. These findings underscore the critical role of the skipped exon in *MYH11*.

Mining differential RNA modifications in CRC

Through sequencing of five paired tissues, the differential RNA modification profile in CRC patients displayed a widespread distribution across the genome (Fig. 4A, Table S2). The distribution of 5-mers, where the proportion of C bases before G modification was lower than other bases, was not perfectly uniform (Fig. 4B). This suggests that the detected differential signal sites exhibit specific base preferences, indicating they are likely non-random biological events.

After filtering differential RNA modification sites by |diff mod rate | > 0.15, these modifications could be categorized into significantly up or down-regulated sites (Fig. 4C). The distribution of RNA site modification rates indicates a tendency towards a normal distribution (Fig. 4D). Exons showed the highest enrichment of differential modification sites, followed by 3'UTRs, with log2 Ratios (obs/exp) of 5.387 and 4.925, respectively. The third most enriched region was 5'UTRs (Fig. 4E), which differential modification sites enriched top regions are closely associated with gene function. The last two regions were intron and intergenic, likely due to Nanopore direct RNA sequencing depended on poly(A) tails, biases towards detecting mature RNAs.

Mining regulation pairs between differential RNA modifications and aberrant ASEs

The reciprocal regulatory relationships between RNA modification and alternative splicing were being investigated. In previous studies, limited researches have delved into the regulatory interactions between RNA modifications and ASEs due to technological constraints. Therefore, in our research, we are pioneering the exploration of these regulatory interactions using a single sequencing technology that is not influenced by other technologies. Utilizing the analysis pipeline, we have identified regulatory pairs of RNA modification and ASEs (Fig. 5A). For gene *MYH11*, two distinct differential RNA modification sites were identified that likely regulate its ASEs (Fig. 5B). The comprehensive analysis has revealed a total of 304 cis-regulation pairs (Table S3). Upon narrowing our focus to cis-regulation pairs, we observed that the distribution of the 5-mer base does not exhibit a specific trend (Fig. 5C). The absence of a dominant sequence motif pattern strengthens the possibility that non-linear features, such as RNA secondary or tertiary structure, mediate the functional coupling for a subset of our identified pairs. Furthermore, the differential modification rate of these cis-regulation pairs' distribution continues to display a propensity towards a normal distribution (Fig. 5D). The distribution of up-regulated modification sites in cis-regulation pairs totals 111, while down-regulated sites amount to 122 (Fig. 5E), with an even distribution observed. Interestingly, the relative gene position of these sites is predominantly concentrated at the beginning of the gene (Fig. 5F). The identification of these potential regulatory pairs is crucial for shedding light on additional research about RNA modifications within tumors.

RNA modification related enzymes and splicing related proteins in CRC

We employed a suite of RNA modification associated enzymes, including m5C and m6A regulators (readers, writers, and erasers), as well as splicing-related proteins such as hnRNP and SR family members, to identify potential regulatory relationships between differential RNA modification sites and aberrant ASEs. This approach was taken to investigate the potential functional relevance of these regulatory pairs.

The expression patterns of m6A-related enzymes are shown in Fig. 6A and mC-related enzymes in Figure S1. Analysis of CRC samples identified two EIF3A alternative splicing transcripts, EIF3A-201 and EIF3A-203, which display distinct profiles of putative modification sites. Based on predicted regulatory interactions, seven transcripts associated with putative RNA modifications were found to be linked to modification sites in EIF3A-201 (Fig. 6B, Table S5). Among these, three genes—NACA (nascent polypeptide-associated complex subunit alpha), LDB1 (LIM domain-binding 1), and PPP3CB (protein

phosphatase 3 catalytic subunit beta)—have been previously implicated in cancer pathogenesis (Table S6). Functional annotation indicated associations of NACA with extracellular exosome processes, LDB1 with Wnt signaling and transcriptional regulation in cancer, and PPP3CB with T-cell receptor and NFAT signaling pathways.

Splicing-related protein expression is summarized in Fig. 6C. Two alternative splicing isoforms, SRSF1-206 and SRSF6-202, were detected in CRC. Consistent with the earlier approach, putative RNA modification-associated transcripts were identified for each: seven corresponding to SRSF6-202 and eighteen to SRSF1-206 (Fig. 6D, Table S5). Several of these transcripts, highlighted in red in the network, represent cancer-functional genes involved in Wnt and Notch signaling pathways, as well as in cellular processes such as proliferation, apoptosis, and migration (Table S6).

Discussion

Long read sequencing can help detect aberrant ASEs. In our research, we identified a total of 263 aberrant ASEs, the majority of which were enriched in critical cancer-related pathways, including the Wnt signaling pathway. Within the interaction network, the ribosomal gene *RPS24* emerged as the top-degree node, corroborating previous research that linked alternative splicing of *RPS24* to cancer [32]. Other prominent nodes, such as *GSK3B*, *FN1*, and *MYH11*, were also identified; *GSK3B* is known to be associated with tumor cell survival and proliferation in CRC [24], while downregulation of *FN1* has been shown to inhibit tumor cell proliferation, migration, and invasion [29]. Particularly noteworthy is the cancer-related gene *MYH11*, for which the proportion of main transcripts underwent significant alterations in CRC. The skipped exon in the MYH11-203 transcript was closely related to domains of other key cancer genes and the analysis indicated that it may serve as a potential diagnostic candidate, subject to further clinical validation.

Our structural prediction suggests that alternative splicing of a single exon may induce a rotational shift in the transmembrane domain of isoform MYH11-203. This conformational change is predicted to potentially influence functional consequences. This rotation remodels the interaction interface, enabling each isoform to engage distinct downstream effectors and ultimately adopt divergent signaling outputs and functional roles. Furthermore, the altered TMD geometry may affect subcellular localization by modulating integration into lipid microdomains or recognition by cellular trafficking machinery. The model illustrates how alternative splicing might fine-tune protein function not only through linear sequence changes but also via allosteric structural rearrangements, providing a sophisticated mechanism for functional divergence from a single gene. This finding draws new attention to aberrant ASEs in cancer, particularly exon skipping, and informs future drug design efforts targeting these critical regions.

RNA modifications have often been overlooked in most studies, primarily due to the difficulty of integrating these modifications with gene expression. In our study, we utilized direct RNA sequencing to simultaneously quantify RNA modifications and gene expression. The level of RNA modification was quantified as a modification rate, which was transformed into an integer matrix. By employing linear regression incorporating covariates such as age and gender, we identified cis-regulatory pairs and trans-regulatory pairs. Notably, multiple trans-regulatory pairs were detected, but focusing solely on the relationships between modification rates and the expression of ASEs is insufficient. To ensure their accuracy, it is essential to consider spatial interaction information among these pairs.

Nanopore direct RNA sequencing enables researchers to uncover various types of differential RNA modifications in CRC. During the analysis, we observed that different types exhibited specific base preferences in their flanking sequences (Fig. 4B). Crucially, by integrating these sites with the expression of specific m6A and m5C regulators, we were able to characterize potential regulatory networks (Fig. 6). This approach indicated that the majority of these putative RNA modification sites were enriched in gene exons. The characterization of binding motifs for methyltransferases, demethylases, and m6A readers has significantly advanced this field. To date, more than 150 distinct RNA modifications have been identified in cellular RNAs, necessitating simultaneous investigations of individual motifs. These findings suggest that there are significant differences in RNA modifications between CRC and normal tissues, providing valuable insights into the RNA modification mechanism underlying this disease. Further research in this area could help in the development of targeted therapies for CRC.

The lack of a dominant linear motif among the RNA modification-ASE pairs suggests that their regulation may not rely solely on sequence codes (Fig. 5C). Instead, this finding points to the involvement of non-linear features, such as RNA structural elements. Modifications like m6A can alter local RNA folding, thereby creating or masking binding sites for splicing regulators. This structure-mediated mechanism provides a plausible explanation for the observed complexity and context-specificity of the interactions, representing a crucial direction for future validation using advanced structural probing techniques.

Based on our systematic screening of regulatory pairs linking RNA modification enzymes (m5C/m6A regulators) and splicing factors (hnRNPs/SR proteins) with

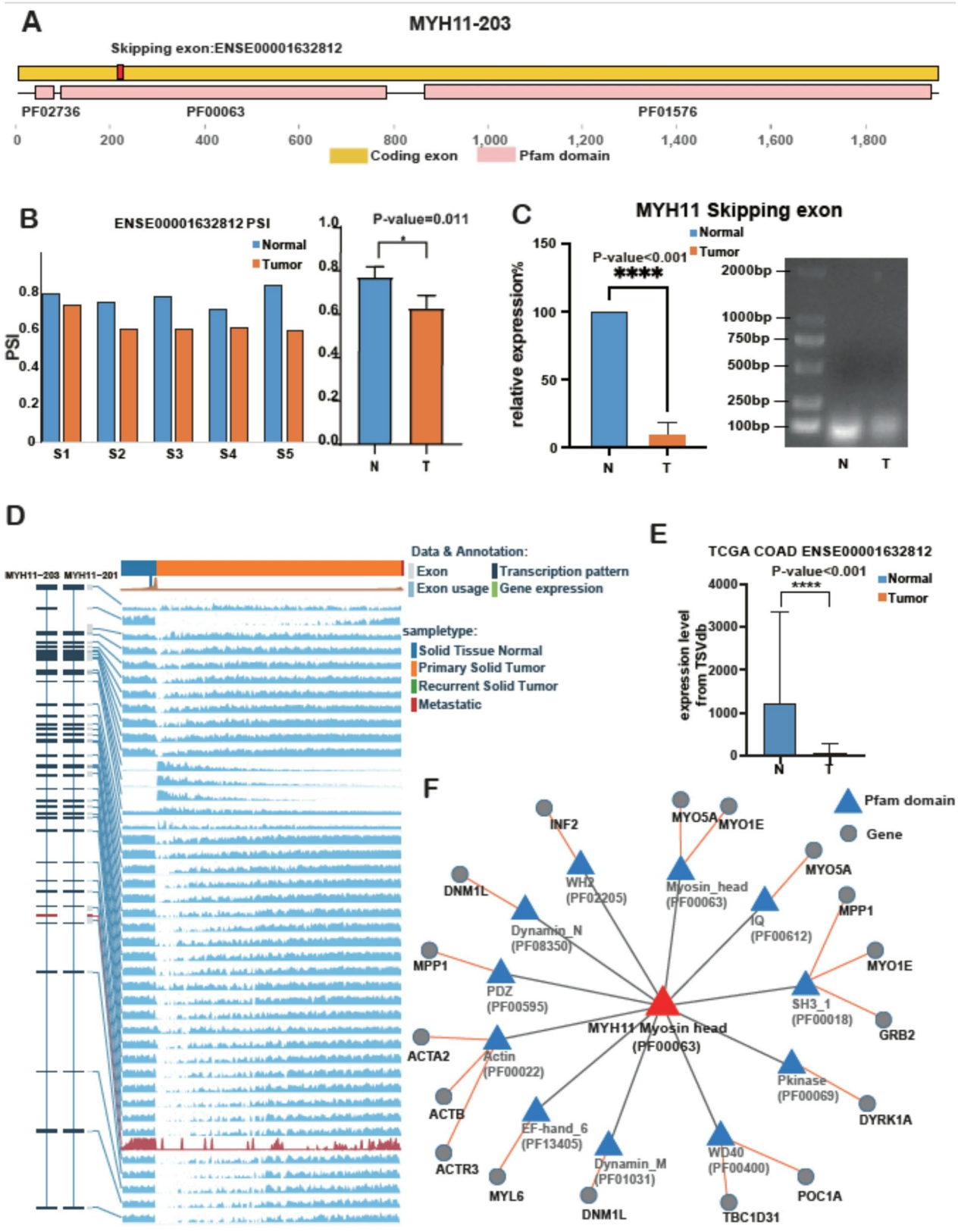


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Aberrant Skipped exon in MYH11-203 **A**. The skipped exon ENSE00001632812 in MYH11-203. The protein domain was in pink and skipped exon in red **B**. PSI values of ENSE00001632812 in paired tissues, with blue bars representing normal tissues and orange bars representing tumor tissues **C**. The ASE in MYH11 were detected by RT-PCR **D**. The exon usage of MYH11 in TCGA COAD cohort. The red exon is ENSE00001632812 **E**. The expression level of ENSE00001632812 from TSVdb significantly different between normal and tumor tissues **F**. The PF00063-related domain-domain interaction network. The blue triangles represent Pfam domains and grey circles are genes

alternative splicing transcripts, we have characterized potential associations between post-transcriptional regulation layers in CRC. The identified relationships, such as those between modifications in EIF3A-201 and transcripts of NACA and LDB1, connect specific splicing events to key cancer pathways, including Wnt and immune signaling. Similarly, the extensive network of transcripts linked to RNA modifications in SRSF1-206 and SRSF6-202 highlights a coordinated mechanism potentially governing oncogenic processes like proliferation and apoptosis. These findings strongly suggest that the interplay between differential RNA modifications and aberrant splicing is not merely coincidental but may represent a critical functional axis contributing to CRC pathogenesis. The identified regulatory pairs provide a valuable resource for understanding the mechanistic basis of RNA dysregulation in tumors and for identifying potential therapeutic targets.

We acknowledge several limitations in our study, primarily concerning the sample size of the discovery cohort. As a pilot investigation utilizing Nanopore direct RNA sequencing, our analysis was restricted to five paired CRC tissues. While this sample size allowed for the detailed characterization of complex splicing isoforms and RNA modifications, it inherently limits statistical power and may not fully capture the extensive inter-patient heterogeneity characteristic of CRC. Furthermore, we acknowledge the lack of extensive experimental validation. This limitation stems from the specific requirements of Nanopore direct RNA sequencing, which demands high input amounts of native poly(A) + RNA. Following the discovery phase, the available high-quality biological material from these clinical samples was almost exhausted, precluding further wet-lab assays. Consequently, this study is positioned as a computational and exploratory analysis. Some identified events might be specific to this cohort, and subtle splicing changes could have been overlooked. To mitigate the risk of false positives associated with a small sample size, we applied stringent filtering criteria, including a biological effect threshold and gene-level statistical correction. Furthermore, we prioritized the external validation of our key finding, the MYH11 exon skipping event, using the large-scale TCGA dataset, which confirmed its robustness. Furthermore, while AlphaFold3 provided structural insights, we did not perform in vitro functional assays to experimentally validate the biological consequences of the MYH11 exon skipping. Therefore, the downstream effects on tumor progression described herein remain hypothetical and warrant future experimental verification. Nevertheless, the other identified ASEs and regulatory pairs should be regarded as preliminary candidates. Future studies involving larger, multi-center clinical cohorts are essential to independently validate these targets and rigorously establish their diagnostic or therapeutic utility before clinical translation.

Materials and methods

Patients and samples

CRC patients were recruited from Beijing Hospital. All patients had newly diagnosed stage II and III CRC within one year prior to recruitment, with histological confirmation and no prior treatment with chemotherapy, immunotherapy, or radiotherapy. Paired tumor and tumor-adjacent normal tissues from five patients were obtained from the surgically removed at Beijing Hospital. All the Patients were microsatellite stability and there were no restrictions on age, gender, or smoking status in this small pilot study. Fresh tissues were collected immediately and stored in liquid nitrogen before testing and all samples were examined by experienced pathologists.

Nanopore direct RNA sequencing and data quality control

Total RNA was isolated using the TRIzol Reagent (Thermo Fisher Scientific) following the manufacturer's instructions for steps of RNA precipitation, washing, and solubilization. The quality of RNA was checked by Capillary Electrophoresis System and at least 300ng of poly(A) tailed RNA was reserved for Oxford Nanopore Direct RNA Sequencing. The RNA library preparation was constructed following the manufacturer's instructions (version: DRS_9195_v4_revB_20Sep2023). The Oxford Nanopore sequencing platform PromethION with an FLO-PRO004RA (Oxford Nanopore Technology) was utilized for RNA library sequencing. The EXP-WSH004 flow cell wash kit (Oxford Nanopore Technology) and the SQK-RNA004 direct RNA Sequencing Kit (Oxford Nanopore Technology) were used for direct RNA sequencing. After the library samples passed quality checks, they were added to the chip.

The raw electronic signals were collected by MinKNOW software (v23.11.8) [41]. To ensure high-quality data generation, basecalling was performed using Dorado (v0.7.2) with the high-accuracy model. We implemented a strict quality control workflow where only reads with a mean Quality Score > 9 were retained to minimize

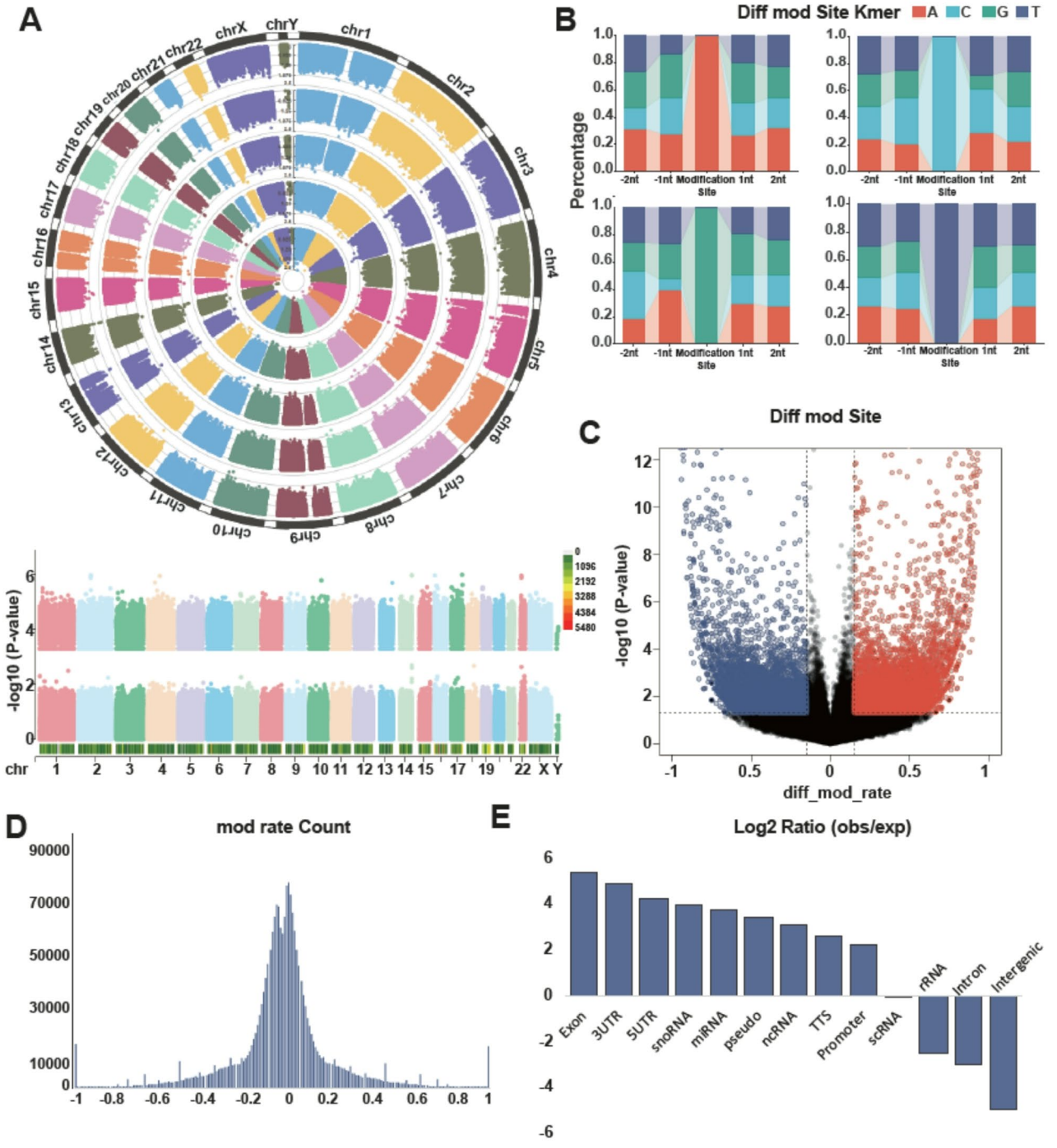


Fig. 4 Differential RNA modification sites in CRCA. Differential modification rate between five paired tumor and normal tissues. The top part of the figure shows the differential modification rates, with each circle representing a sample and the sites of differential RNA modifications indicated. The bottom part illustrates differential RNA modifications in one paired tissue. The intensity of red color indicates a great tendency for modifications exist in these regions. Only | diff mod rate| > 0.15 were present B. The distribution of base in 5-mer, where modified base sits was in the middle if modified. The four subgraphs correspond to the four bases C. The volcano plot of modification sites in five samples. The sites with |diff mod rate| > 0.15 were defined as differential modification sites. The up-regulated modification sites were red D. The count distribution of RNA site modification rates E. The normalized proportion of differential modification sites distribution in human genome, which is the observed distribution proportion divide by the expected distribution proportion. The higher ratio indicates a greater tendency for modification sites exist in these elements

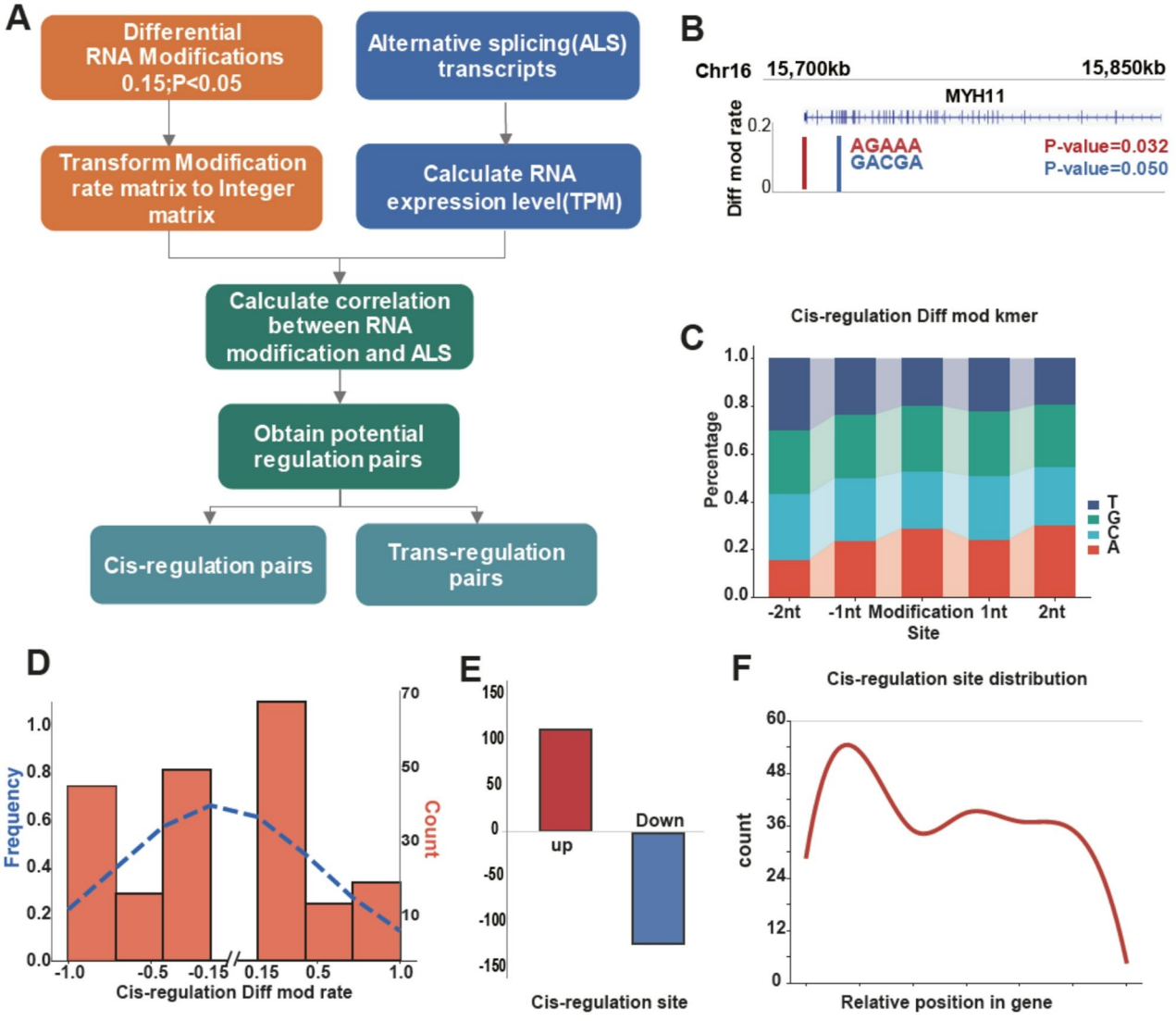


Fig. 5 Exploring the regulatory relationship between differential RNA modifications and aberrant ASEs in CRC **A**. Analysis pipeline for identifying regulated pairs. Cis-regulation pairs involve differential RNA modification sites and ASEs within the same gene **B**. Differential RNA modification sites that significantly regulate ASEs in MYH11 gene **C**. Distribution of cis-regulation within the 5-mer base sequence, where modified base sits were in the middle if modified **D**. Distribution of differential modification rates at cis-regulation sites. The orange bar represents the count number and the blue line shows the frequency fitting curve **E**. Number of up and down regulated modification sites in cis-regulation pairs **F**. Relative distribution of cis-regulation sites within the gene. The x-axis represents the relative gene position, with the locus closer to the start site when the modification site is in proximity to the gene start

sequencing error rates, consistent with the requirements of the SQK-RNA004 chemistry. Notably, as direct RNA sequencing is a PCR-free method that sequences native RNA molecules, identical reads were considered genuine biological duplicates rather than PCR artifacts; therefore, computational duplicate removal was not applied to preserve the accurate quantification of transcript abundance.

Aberrant ASEs in CRC

Differential splicing analysis was performed using SUPPA2 [49] with the empirical method. To ensure the reliability of the identified events, we applied a stringent filtering strategy. The PSI metric was computed to indicate relative abundances, while ΔPSI, calculated as PSI tumor minus PSI normal, provided insights into splicing alterations between tumor and normal samples. First, transcripts with low expression were excluded using a default abundance threshold (TPM > 1.0). Second, significant splicing events were defined based on a dual criterion combining biological effect size and statistical significance: (1) an absolute change in PSI ($|\Delta PSI| > 0.05$), implemented via the --lower-bound parameter to capture biologically relevant alterations; and (2) a significance threshold of P-value < 0.05. Notably, we utilized the gene-correction feature in SUPPA2 to correct for multiple testing arising from multiple events within the same

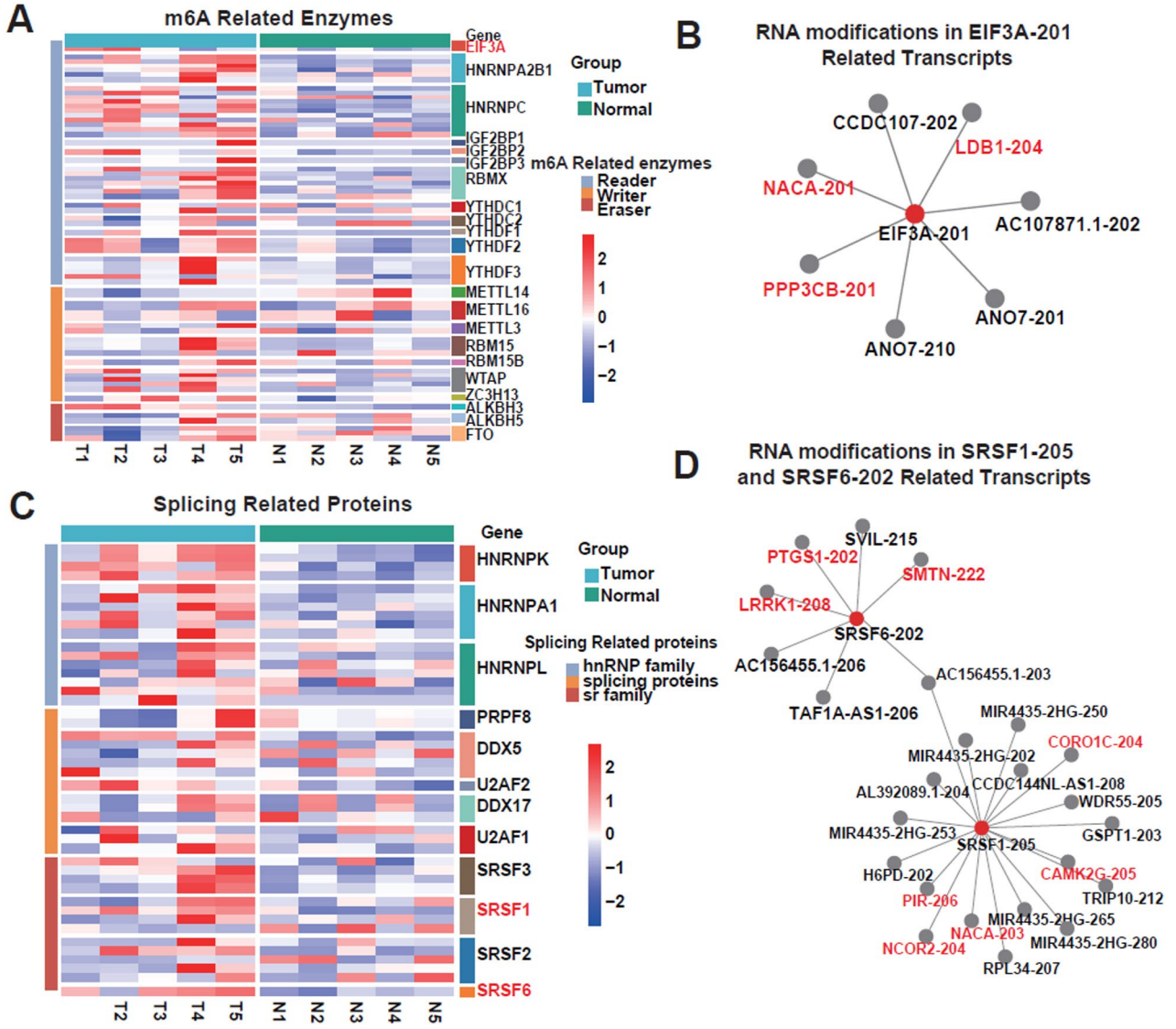


Fig. 6 Analysis of RNA modification-related enzymes and splicing-related proteins in CRC **A**. The expression of m6A-related enzymes in CRC. Heatmap displays expression levels of key m6A regulators (writers, erasers, readers) in tumor versus normal tissues. Gene EIF3A shows distinct transcript alternative splicing patterns **B**. Regulatory network of EIF3A-201 modification associated transcripts. Network diagram highlights seven RNA modification-linked transcripts regulated by EIF3A-201, including cancer-related genes NACA, LDB1, and PPP3CB (red nodes) **C**. Expression profiling of splicing-related proteins. Heatmap shows expression of hnRNP, SR family and other proteins in tumor and normal samples. SRSF6-202 and SRSF1-206 show distinct transcript alternative splicing patterns **D**. Splicing factor-regulated transcript networks. Networks depict RNA modification-associated transcripts for SRSF6-202 (7 transcripts) and SRSF1-206 (18 transcripts). Cancer-functional genes such as NOCR2 and CAMK2G are involved in signaling pathways and cellular processes

gene, thereby reducing false positives associated with transcript-level variability. The exon usage of MYH11 in TCGA COAD dataset was downloaded from TSVdb [42].

Visualization of ASEs was carried out using the R package pheatmap (https://github.com/raivokolde/pheatmap), while enrichment analysis was conducted using the David [43] and protein-protein interaction (PPI) and domain-domain interaction (DDI) networks were explored using Diggger [44]. Gene and transcript quantification was performed using Featurecount [45] and Nanocount [46], enhancing the comprehensive analysis of splicing dynamics in CRC. The three-dimensional structures of protein isoforms using AlphaFold3 [47] and subsequently performed a structural alignment with TM-align [48].

RNA extraction and RT-qPCR

Total RNA was isolated from the tissue using TRIzol reagent. Reverse transcription was performed using the FastKing one-step de-genomic cDNA first strand synthesis premix reagent according to the manufacturer's instructions. Quantitative real-time PCR (qRT-PCR) was

carried out on an Applied Biosystems™ QuantStudio™ 5 using PowerUp SYBR Green Master Mix. Human β-actin was used as an internal control to normalize the relative expression levels. The primer sequences used were listed in Table S4. The data was expressed as the mean ± SD. The differences between two groups were analyzed using the unpaired *t*-test.

Differential RNA modification profiling in CRC

After obtaining the pod5 files, the pod5_convert_to_fast5.py script (https://github.com/nanoporetech/pod5-file-format) was utilized to convert the pod5 file format into fast5 files to meet the requirements of other tools. Subsequently, Dorado software (v0.7.2) (https://github.com/nanoporetech/dorado) was employed for baseca lling from the pod5 files. The xPore (v2.0) [49], including dataprep, diffmod, and postprocessing modules, was then utilized to identify and analyze differential RNA modification sites. It is important to note that xPore detects significant deviations in electrical current signals, which indicate the presence of differential modifications but do not chemically identify the specific modification type. Consequently, these events are referred to as differential RNA modification sites, and their potential biological identities were inferred based on their regulatory associations with specific modification enzymes. For mapping precision, clean reads were aligned to the Gencode v35 reference genome using minimap2 (v2.17-r941) [50] with the splice-aware preset (-ax splice) to effectively handle long-read spanning junctions, while secondary and supplementary alignments were filtered using samtools (v1.11) [51] to ensure mapping specificity. To further reduce noise, Nanopolish (v0.13.2) [52] was utilized to perform "resquiggling", aligning raw electrical events to the reference sequence to improve resolution for modification detection. The analysis was guided by the optimized cutoffs at sequencing coverage depths [53], including a threshold of |diff_mod_rate_KO_vs_WT|>0.15. Visualization of results involved the utilization of the R packages CMplot [54] and SRplot [55] for visualizing chromosome distribution and displaying 5-mer bases. Furthermore, Homer (v4.11.1) [56] annotatePeaks.pl was employed for the annotation analysis phase.

Combine mining RNA modification and ASE regulation pairs in CRC

Like xQTL studies, we integrated RNA modification matrix with ASEs to explore the interplay between these two regulatory mechanisms. For analysis, we employed the R package MatrixEQTL [57] for fast combination analysis by using linear regression with either additive or ANOVA genotype effects.

To enable efficient regression analysis, we converted the modification rate matrix into an integer matrix, with patient gender and age included as covariates in the regression models. The relative position within a gene was determined by dividing the gene's position by its length, allowing for a comprehensive understanding of the distribution of regulatory elements in the gene.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40246-026-00931-0.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

Acknowledgements

The authors would like to thank Clinical Biobank, Beijing Hospital for the sample storage and data transmission.

Author contributions

L.Z. and H.C. wrote the main manuscript text. J.Cui and J.S. prepared figures. T.S., Y.L., J.Cai, Z.C., G.S., H.L., S.X., and X.T. contributed to data curation and analysis. F.X. and G.Z. supervised the project. All authors reviewed and approved the final manuscript.

Funding

This work was supported by the National High Level Hospital Clinical Research Funding (BJ-2024-153, BJ-2024-095), CAMS Innovation Fund for Medical Sciences (grant no. 2021-I2M-1-050), the Fundamental Research Funds for the Central Universities (NO: 3332019121) and the Special Research Fund for Central Universities, Peking Union Medical College (NO: 3332023090).

Data availability

The sequencing datasets generated and analysed during the current study are available in the GSA-Human repository, with the accession number HRA008180.

Declarations

Ethics approval and consent to participate

The research was approved by Beijing Hospital Ethics Committee(2024BJYYEC-KY146-01). This study has been approved by the Ethics Committee of Beijing Hospital and Human Genetic Resource Administration of China (HGRAC), and all procedures were conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants before testing.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Clinical Biobank, Beijing Hospital, National Center of Gerontology; Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing 100730, China

[2]Department of General Surgery, Beijing Hospital, National Center of Gerontology; Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing 100730, China
[3]Department of Clinical Laboratory, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China

Received: 27 October 2025 / Accepted: 3 February 2026
Published online: 13 February 2026

References

1. Marasco LE, Kornblihtt AR. The physiology of alternative splicing. Nat Rev Mol Cell Biol. 2023;24:242–54.
2. Chen Y, Huang M, Liu X, Huang Y, Liu C, Zhu J, et al. Alternative splicing of mRNA in colorectal cancer: new strategies for tumor diagnosis and treatment. Cell Death Dis. 2021;12:752.
3. Hu Y, Liu W, Qiang D, Huang K, Ruan Q, Chen L. Emerging roles of RNA-binding protein HnRNPM in alternative splicing regulation and UTMD mediated sh-hnRNPM/CMBs suppress the proliferation of colorectal cancer. Oncogene. 2025;44:4549–63.
4. López-Sánchez LM, Jiménez-Izquierdo R, Peñarando J, Mena R, Guil-Luna S, Toledano M, et al. SWATH-based proteomics reveals processes associated with immune evasion and metastasis in poor prognosis colorectal tumours. J Cell Mol Med. 2019;23:8219–32.
5. Zhang X, Li M, Ye S, Shen K, Yuan H, Bakhtyar S, et al. CBD2: A functional biomarker database for colorectal cancer. Imeta. 2024;3:e155.
6. Nguyen TB, Miramontes R, Chillon-Marinas C, Maimon R, Vazquez-Sanchez S, Lau AL, et al. Aberrant splicing in huntington's disease accompanies disrupted TDP-43 activity and altered m6A RNA modification. Nat Neurosci. 2025;28:280–92.
7. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA modifications in gene expression regulation. Cell. 2017;169:1187–200.
8. Li S, Mason CE. The pivotal regulatory landscape of RNA modifications. Annu Rev Genomics Hum Genet. 2014;15:127–50.
9. Tang Q, Li L, Wang Y, Wu P, Hou X, Ouyang J, et al. RNA modifications in cancer. Br J Cancer. 2023;129:204–21.
10. Barbieri I, Kouzarides T. Role of RNA modifications in cancer. Nat Rev Cancer. 2020;20:303–22.
11. Schubert C. Technology Feature | Epitranscriptomics: RNA revisited. Science 2019;364.
12. Wang F, Wang Q, Wu Y, Huang Z, Zhong X, Wang H, et al. CircMVP promotes METTL3 activation mediated CTNNB1 m6A modification in the Inhibition of colorectal cancer in B7-H3 dependence antitumor immunity. Int J Biol Sci. 2025;21:306–27.
13. Zhang L, Mao Z, Yin K, Wang S. Review of METTL3 in colorectal cancer: from mechanisms to the therapeutic potential. Int J Biol Macromol. 2024;277:134212.
14. Chauhan N, Saxena K, Jain U. Single molecule detection; from microscopy to sensors. Int J Biol Macromol. 2022;209:1389–401.
15. Kim Y, Saville L, O'Neill K, Garant JM, Liu Y, Haile-Merhu S, et al. Nanopore direct RNA sequencing of human transcriptomes reveals the complexity of mRNA modifications and crosstalk between regulatory features. Cell Genom. 2025;5:100872.
16. Zhu ZM, Huo FC, Zhang J, Shan HJ, Pei DS. Crosstalk between m6A modification and alternative splicing during cancer progression. Clin Transl Med. 2023;13:e1460.
17. Widagdo J, Anggono V, Wong JJ. The multifaceted effects of YTHDC1-mediated nuclear m(6)A recognition. Trends Genet. 2022;38:325–32.
18. Azzam SK, Alsafar H, Sajini AA. FTO m6A demethylase in obesity and cancer: implications and underlying molecular mechanisms. Int J Mol Sci 2022;23.
19. Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF, et al. Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. Mol Cell. 2016;61:507–19.
20. Tian C, Huang Y, Li Q, Feng Z, Xu Q. Mettl3 regulates osteogenic differentiation and alternative splicing of Vegfa in bone marrow mesenchymal stem cells. Int J Mol Sci 2019;20.
21. Xu RY, Ding Z, Zhao Q, Ke TY, Chen S, Wang XY et al. An alternatively spliced variant of METTL3 mediates tumor suppression in hepatocellular carcinoma. Genes (Basel) 2022;13.
22. Kim MK, McGarry TJ, Flatow POB, Golden JM, Licht AA. An integrated genome screen identifies the Wnt signaling pathway as a major target of WT1. Proc Natl Acad Sci U S A. 2009;106:11154–9.
23. García-Ibarbia C, Delgado-Calle J, Casafont I, Velasco J, Arozamena J, Pérez-Núñez MI, et al. Contribution of genetic and epigenetic mechanisms to Wnt pathway activity in prevalent skeletal disorders. Gene. 2013;532:165–72.
24. Shakoori A, Ougolkov A, Yu ZW, Zhang B, Modarressi MH, Billadeau DD, et al. Deregulated GSK3beta activity in colorectal cancer: its association with tumor cell survival and proliferation. Biochem Biophys Res Commun. 2005;334:1365–73.
25. Katoh M, Katoh M. WNT signaling and cancer stemness. Essays Biochem. 2022;66:319–31.
26. Guo M, Zhang X, Wang G, Sun J, Jiang Z, Khadarian K, et al. miR-603 promotes glioma cell growth via Wnt/β-catenin pathway by inhibiting WIF1 and CTNNBIP1. Cancer Lett. 2015;360:76–86.
27. Li J, Tian L, Jing Z, Guo Z, Nan P, Liu F, et al. Cytoplasmic RAD23B interacts with CORO1C to synergistically promote colorectal cancer progression and metastasis. Cancer Lett. 2021;516:13–27.
28. Ferlizza E, Solmi R, Miglio R, Nardi E, Mattei G, Sgarzi M, et al. Colorectal cancer screening: assessment of CEACAM6, LGALS4, TSPAN8 and COL1A2 as blood markers in faecal immunochemical test negative subjects. J Adv Res. 2020;24:99–107.
29. Cai X, Liu C, Zhang TN, Zhu YW, Dong X, Xue P. Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. J Cell Biochem. 2018;119:4717–28.
30. Liang W, Wu J, Qiu X. LINC01116 facilitates colorectal cancer cell proliferation and angiogenesis through targeting EZH2-regulated TPM1. J Transl Med. 2021;19:45.
31. Wang RJ, Wu P, Cai GX, Wang ZM, Xu Y, Peng JJ, et al. Down-regulated MYH11 expression correlates with poor prognosis in stage II and III colorectal cancer. Asian Pac J Cancer Prev. 2014;15:7223–8.
32. Park J, Nam DH, Kim D, Chung YJ. RPS24 alternative splicing is a marker of cancer progression and epithelial-mesenchymal transition. Sci Rep. 2024;14:13246.
33. Gorący J, Bogacz A, Uzar I, Wolek M, Łochyńska M, Ziętek P et al. The analysis of NADPH Quinone reductase 1 (NQO1) polymorphism in Polish patients with colorectal cancer. Biomolecules 2021;11.
34. Ding Y, Lv Z, Cao W, Shi W, He Q, Gao K. Phosphorylation of INF2 by AMPK promotes mitochondrial fission and oncogenic function in endometrial cancer. Cell Death Dis. 2024;15:65.
35. Ji YJ, Shao Y, Zhang J, Zhang X, Qiang P. Bromodomain-containing protein 4 Silencing by microRNA-765 produces anti-ovarian cancer cell activity. Aging. 2021;13:8214–27.
36. Ng CW, Tsang YTM, Gershenson DM, Wong KK. The prognostic value of MEK pathway-associated Estrogen receptor signaling activity for female cancers. Br J Cancer. 2024;130:1875–84.
37. Jusue-Torres I, Tiv R, Ricarte-Filho JC, Mallisetty A, Contreras-Vargas L, Godoy-Calderon MJ, et al. Myo1e overexpression in lung adenocarcinoma is associated with increased risk of mortality. Sci Rep. 2023;13:4107.
38. Hu H, Zhang S, Xiong S, Hu B, He Y, Gu Y. ACTR3 promotes cell migration and invasion by inducing epithelial mesenchymal transition in pancreatic ductal adenocarcinoma. J Gastrointest Oncol. 2021;12:2325–33.
39. Guo C, Liu S, Wang J, Sun MZ, Greenaway FT. ACTB in cancer. Clin Chim Acta. 2013;417:39–44.
40. Rammohan M, Harris E, Bhansali RS, Zhao E, Li LS, Crispino JD. The chromosome 21 kinase DYRK1A: emerging roles in cancer biology and potential as a therapeutic target. Oncogene. 2022;41:2003–11.
41. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, et al. Evaluation of Oxford nanopore's minion sequencing device for microbial whole genome sequencing applications. Sci Rep. 2018;8:10931.
42. Sun W, Duan T, Ye P, Chen K, Zhang G, Lai M, et al. TSVdb: a web-tool for TCGA splicing variants analysis. BMC Genomics. 2018;19:405.
43. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
44. Louadi Z, Yuan K, Gress A, Tsoy O, Kalinina OV, Baumbach J, et al. DIGGER: exploring the functional role of alternative splicing in protein interactions. Nucleic Acids Res. 2021;49:D309–18.
45. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
46. Gleeson J, Leger A, Prawer YDJ, Lane TA, Harrison PJ, Haerty W, et al. Accurate expression quantification from nanopore direct RNA sequencing with nanocount. Nucleic Acids Res. 2022;50:e19.

47. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. Nature. 2024;630:493–500.
48. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005;33:2302–9.
49. Pratanwanich PN, Yao F, Chen Y, Koh CWQ, Wan YK, Hendra C, et al. Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. Nat Biotechnol. 2021;39:1394–402.
50. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
51. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al. Twelve years of samtools and BCFtools. Gigascience 2021;10.
52. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de Novo using only nanopore sequencing data. Nat Methods. 2015;12:733–5.
53. Tan L, Guo Z, Wang X, Kim DY, Li R. Utilization of nanopore direct RNA sequencing to analyze viral RNA modifications. mSystems. 2024;9:e0116323.
54. Yin LL, Zhang HH, Tang ZS, Xu JY, Yin D, Zhang ZW, et al. rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-wide association study. Genomics Proteom Bioinf. 2021;19:619–28.
55. Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, et al. SRplot: A free online platform for data visualization and graphing. PLoS ONE. 2023;18:e0294236.
56. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of Lineage-Determining transcription factors prime -Regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89.
57. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28:1353–8.

Publisher's note