

RESEARCH

Open Access



Construction of prognostic predictor by comprehensive analyzing alternative splicing events for colon adenocarcinoma

Yaqi Qu¹, Yujia Chen², Le Zhang³ and Lifei Tian^{1*}

Abstract

Background: Colon adenocarcinoma (COAD) is one of the most common malignant tumors, with high incidence and mortality rates worldwide. Reliable prognostic biomarkers are needed to guide clinical practice.

Methods: Comprehensive gene expression with alternative splicing (AS) profiles for each patient was downloaded using the SpliceSeq database from The Cancer Genome Atlas. Cox regression analysis was conducted to screen for prognostic AS events. The R package limma was used to screen differentially expressed genes (DEGs) between normal and tumor samples in the COAD cohort. A Venn plot analysis was performed between DEGs and prognostic AS events, and the DEGs that co-occurred with prognostic AS events (DEGAS) were identified. The top 30 most-connected DEGAS in protein–protein interaction analysis were identified through Cox proportional hazards regression to establish prognostic models.

Results: In total, 350 patients were included in the study. A total of 22,451 AS events were detected, of which 2004 from 1439 genes were significantly associated with survival time. By overlapping these 1439 genes with 6455 DEGs, 211 DEGs with AS events were identified. After the construction of the protein–protein interaction network, the top 30 hub genes were included in a multivariate analysis. Finally, a risk score based on 12 genes associated with overall survival was established ($P < 0.05$). The area under the curve was 0.782. The risk score was an independent predictor ($P < 0.001$).

Conclusions: By exploring survival-associated AS events, a powerful prognostic predictor consisting of 12 DEGAS was built. This study aims to propose a novel method to provide treatment targets for COAD and guide clinical practice in the future.

Keywords: Colon adenocarcinoma, Alternative splicing, Prognostic risk score, Differentially expressed genes

Introduction

Colorectal cancer (CRC) is the third most common cancer and the third most common cause of tumor-related death. The incidence and mortality rate of CRC have increased dramatically over the past three decades in China because of changes in lifestyle factors [1]. CRC includes colon

adenocarcinoma (COAD) and rectal adenocarcinoma. COAD is one of the main causes of cancer-associated deaths worldwide [2]. Approximately 41% of all CRCs occur in the proximal colon, with approximately 22% involving the distal colon and 28% involving the rectum. The prognosis of COAD is less than satisfactory; over 1.8 million new COAD cases and 881,000 deaths occurred in 2018 [3]. Therefore, there is a need to discover new molecules capable of identifying pre-cancerous lesions.

* Correspondence: awdcv789@163.com

¹Department of General Surgery, Shaanxi Provincial People's Hospital, Xi'an 710068, China

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



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Alternative splicing (AS), a post-transcriptional process that modifies more than 90% of all human genes, contributes significantly to the diversity of the transcriptome and the resultant proteome [4, 5]. Exonic regions of the pre-mRNA are spliced together, and intronic regions are removed to form the final mRNA products. There is increasing evidence that the transcript variants of a gene may have opposing roles and that widespread splicing disorders are one of the molecular markers of tumorigenesis and participate in many oncogenic processes, including proliferation, invasion and metastasis, apoptosis, hypoxia, metabolism, angiogenesis, and immune escape [6, 7].

It is generally believed that dysregulation of gene expression plays an important role in the genesis and development of tumors, and studies in this area have led to improvements in diagnosis and identification of prognostic biomarkers and new therapeutic targets by analyzing the expression level of genes in various types of cancer [8, 9]. Various studies have shown that specific siRNAs, lncRNAs, and miRNAs affect the tumorigenesis and prognosis of COAD [10, 11]. Some gene modifications such as acetylation and N6-methyladenosine are also research hotspots in the field of COAD [12, 13]. However, these studies have largely ignored AS of genes. Therefore, our understanding of functionally important splicing events that contribute to tumorigenesis and the mechanisms that lead to aberrant splicing in COAD is very limited.

In this study, we aimed to explore the role of RNA AS events in the progression of COAD. To provide a deeper insight into the tumor biology, this study further explored the relationship between those molecules and the prognosis of COAD to gain a better understanding of prognostic AS events in COAD.

Methods

Data collection of AS events and data analysis

The Cancer Genome Atlas (TCGA), a large-scale cancer genomic project based on next-generation sequencing and microarrays to discover molecular aberrations at the DNA, RNA, protein, and epigenetic levels, started approximately 10 years ago to discover major genetic alterations in large cohorts of selected tumors, specifically those of brain, lung, and ovarian cancers. The RNA-seq data of the COAD cohort was obtained from the TCGA data portal. The SpliceSeq tool from TCGA was used to obtain the mRNA splicing pattern, known as AS, of COAD samples from the TCGA database. In SpliceSeq, a value known as Percent Spliced In ranging from 0 to 1 was calculated for each detected AS event in a gene. The R package

UpSetR (version 1.3.3) was used to generate an UpSet plot for the quantitative analysis of interactive sets between seven AS types.

Data on the clinical parameters of the COAD cohort were downloaded from TCGA. To rule out the effects of surgery leading to death, patients with incomplete clinical information and a survival time of less than 30 days were excluded from the cohort. Finally, 350 patients with COAD with full clinical follow-up data were included in the study.

Screening for survival-associated AS events in the COAD cohort

Univariate Cox analysis was conducted to determine survival-associated AS events for COAD ($P < 0.05$).

Screening for differentially expressed genes (DEGs) in the COAD cohort

The R package limma was used to search for DEGs between normal and COAD test subjects. The cutoff criteria were an adjusted P value of < 0.05 and a logFC value of ≥ 1 or ≤ -1 . A volcano map was drawn to show the criteria for selecting differential genes using the R package ggplot2. A hierarchical clustering heat map of DEG expression was visualized with the R package pheatmap.

Overlap between survival-associated AS genes and DEGs

The overlap between survival-associated AS genes and DEGs was screened using Funrich (<http://www.funrich.org/>), and Venn diagrams were used to identify DEGs with survival-associated AS. Specific genes that were differentially expressed and occurred with AS events were obtained. We termed these genes as “differentially expressed genes overlapping with AS events (DEGAS)” hereafter.

Functional enrichment of DEGs by gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses

Metascape (<http://metascape.org/gp/index.html#/main/step1>), an online analysis tool suite with the function of Integrated Discovery and Annotation, mainly provides typical batch annotation and GO term enrichment analysis to highlight the most relevant GO terms associated with a given gene. GO and KEGG pathway enrichment analyses were used to identify DEGAS using Metascape. $P < 0.05$ was considered statistically significant.

Construction of prognostic models for the COAD cohort

The top 30 most-connected DEGAS in the protein–protein interaction (PPI) network were identified using Cytoscape (version 3.6.1). Multivariate Cox regression

analysis was used to assess the prognostic value of DEGAS for overall survival (OS). We constructed a prognostic model using the sum of DEGAS expression levels weighted by the regression coefficient (β) from the multivariate analysis:

$$\text{RiskScore} = \sum_{i=1}^n \text{exp}_i * \beta_i$$

where n is the total DEGAS sequence number in the risk score, i is the sequence number of DEGAS in risk score, exp_i is the expression of DEGAS, and β_i is the regression coefficient of gene i in the multivariate Cox regression analysis. The median score was used as a cutoff value to classify patients with COAD into low-risk and high-risk groups.

Survival analyses were assessed using the Kaplan–Meier (KM) method and the two-tailed log-rank test. The efficiency of the predictive model was further assessed using receiver operating characteristic (ROC) curves with the survivalROC package in R. The area under the curve (AUC) of the ROC curves was calculated. All analyses were performed using R software (version 3.6.1).

The risk score was combined with clinical factors using univariate and multivariate Cox regression to identify whether it could be used as an independent prognostic predictor for survival time. The results were represented as forest plots generated by the forestplot R package.

Networks of survival-associated AS genes and splicing factors

The expression of splicing factor genes in the mRNA splicing pathway was investigated by analysis of the level 3 mRNA-seq data in TCGA. Next, the correlation between survival-associated splicing factors and DEGAS was analyzed using the Spearman test. Cytoscape software was used to construct the interaction network. $P < 0.01$ and coefficient = 0.6 were used as the inclusion criteria.

Data analysis and generation of graphs

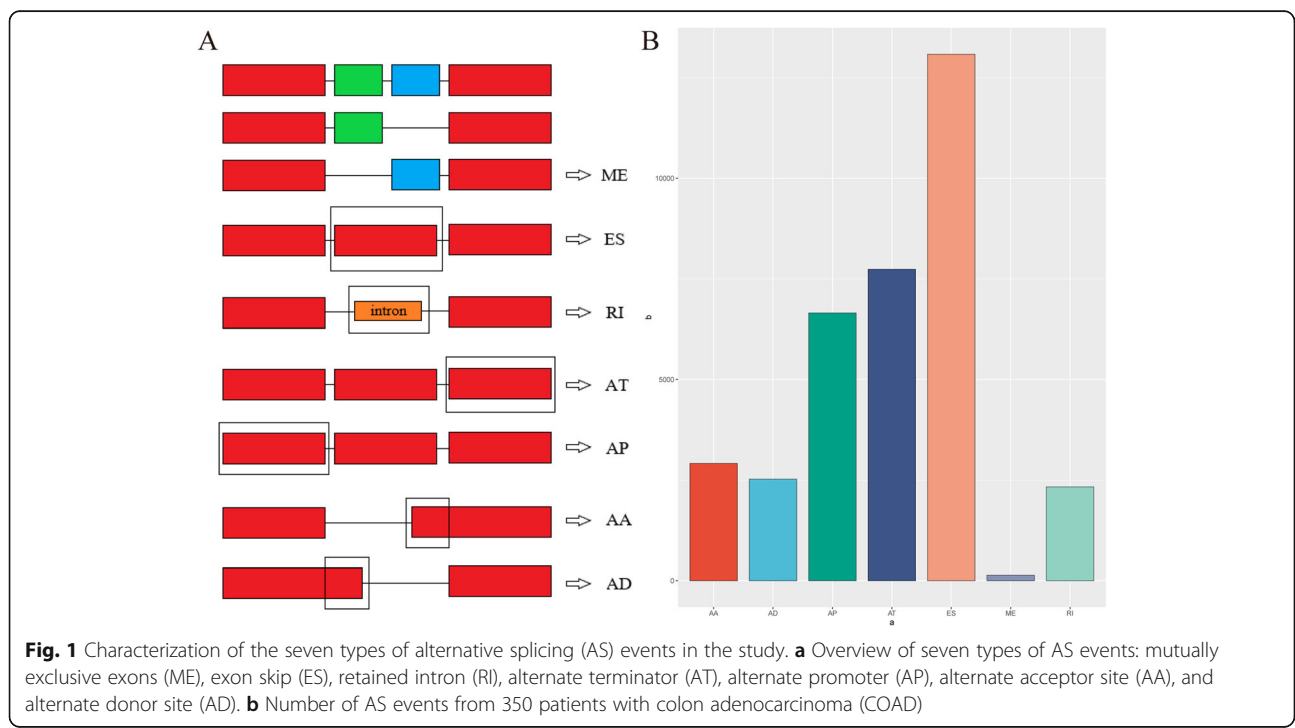
Results

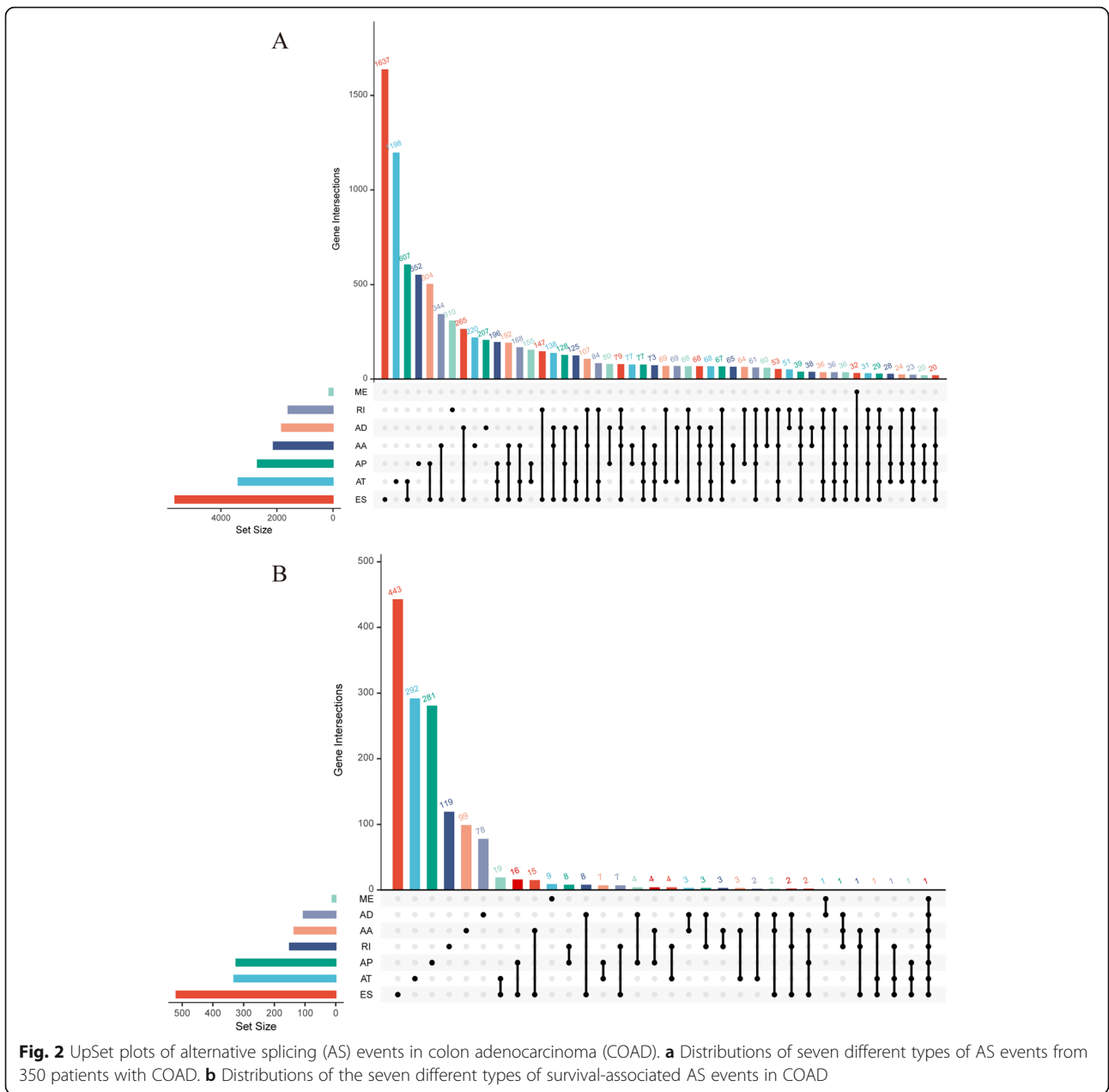
Overview of AS events in COAD

Comprehensive AS events were identified in seven splicing types: exon skip (ES), mutually exclusive exon (ME), retained intron (RI), alternate promoter (AP), alternate terminator (AT), alternate donor site (AD), and alternate acceptor site (AA) (Fig. 1a). Overall, 22,451 AS events were detected, consisting of 74 ME in 74 genes, 1661 RI in 1187 genes, 1682 AD in 1309 genes, 1999 AA in 1574 genes, 4696 AT in 2770 genes, 3972 AP in 2325 genes, and 8367 ES in 4430 genes (Figs. 1b and 2a).

AS events associated with survival time

The results of univariate Cox regression analysis showed that 2004 AS events in 1439 genes were significantly associated with OS ($P < 0.05$), including 10



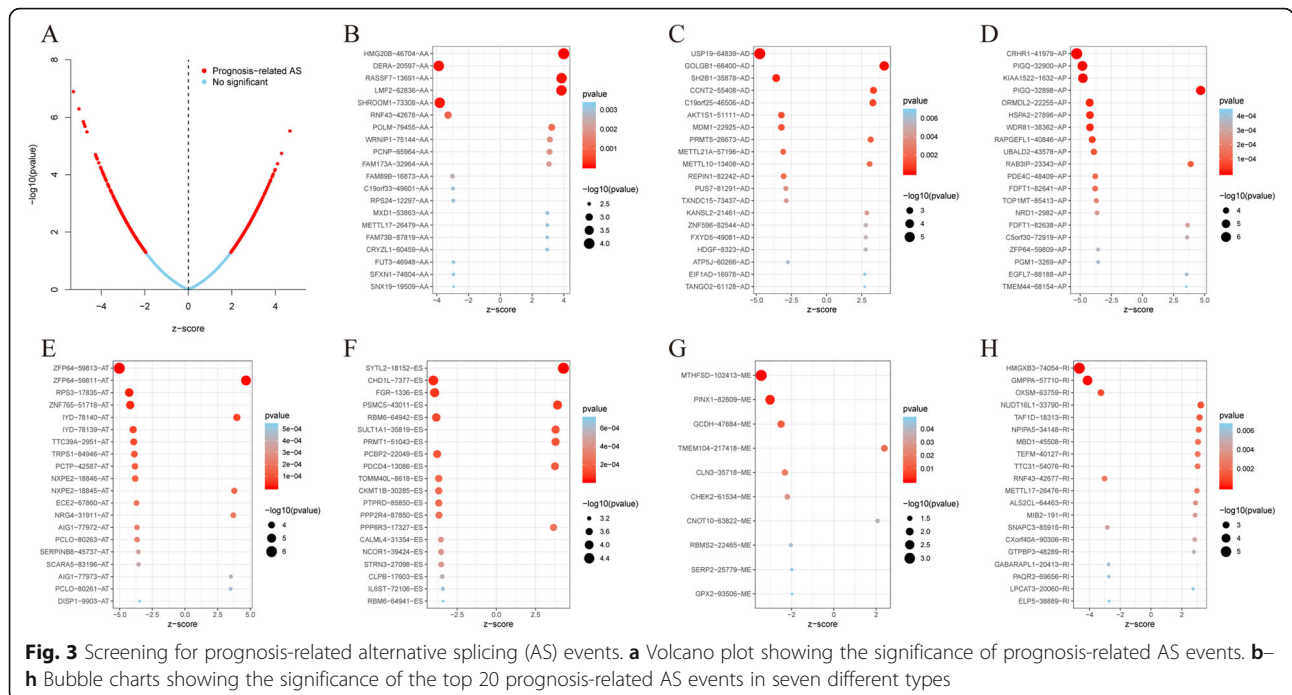


ME in 10 genes, 158 RI in 149 genes, 105 AD in 104 genes, 140 AA in 134 genes, 517 AT in 330 genes, 448 AP in 323 genes, and 586 ES in 518 genes (Fig. 2b). The top 20 of each AS type with the smallest *P* value are presented in Fig. 3. As depicted in the plot, one gene may contain more than one type of AS event.

DEGs related to survival-associated AS events

A total of 6455 DEGs were identified from the TCGA COAD cohort, including 4558 genes that were significantly upregulated and 1897 genes that were

significantly downregulated (Fig. 4; *P* < 0.01, fold-change > 1). To identify DEGAS, we overlapped the 6445 DEGs and 1439 survival-associated AS genes. As a result, 211 DEGs were identified that overlapped with survival-associated AS events. The results are presented as a Venn diagram in Fig. 5a). Principal component analysis (PCA) was performed on the DEGAS to comprehensively characterize them (Fig. 5b–h). Multivariate Cox regression analysis was performed to evaluate the prognostic effects of DEGAS. In COAD, AD, RI, AT, AA, ES, and AP had an AUC > 0.65, and ES had the most satisfactory performance



(AUC = 0.98), while ME had the worst performance (AUC = 0.595; Fig. 5i and j).

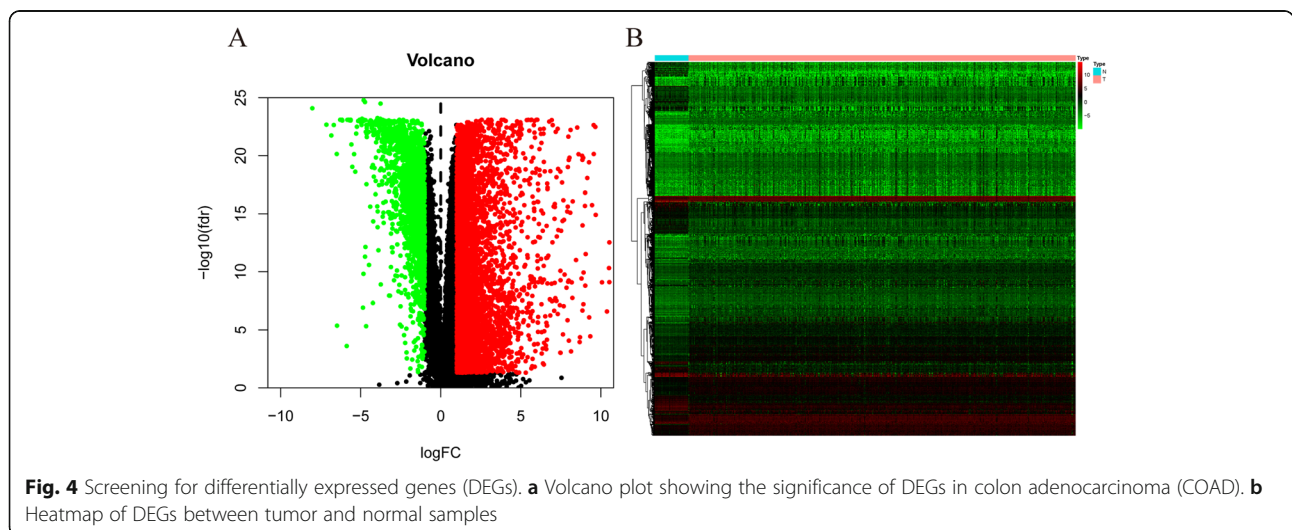
Construction and analysis of PPI network

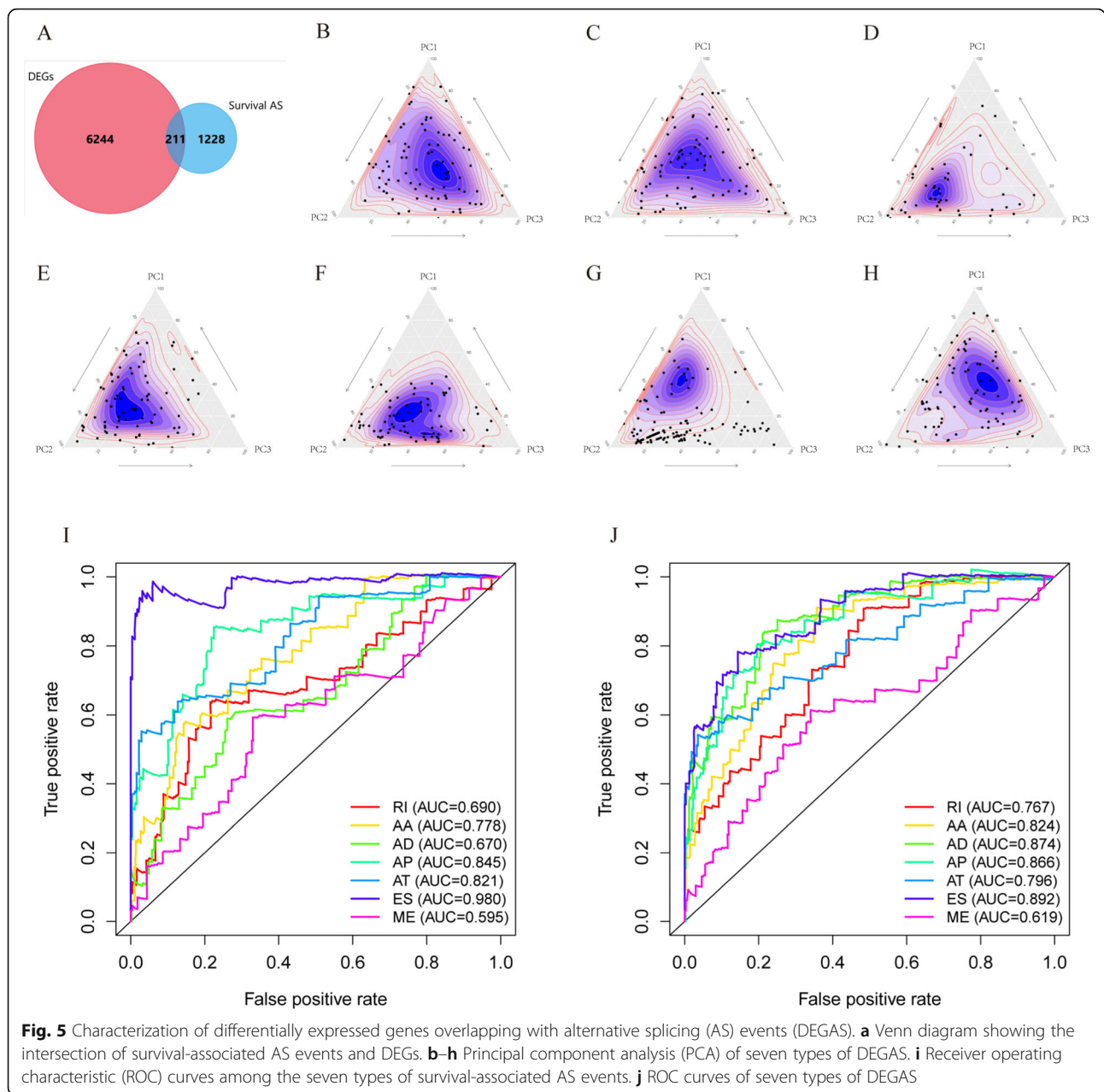
To explore the interactions among DEGAS, a string database was mapped for the most significant prognostic genes with a score > 0.4 using Cytoscape (Fig. 6a). Different types of AS events were shown in different colors. The results indicated that the majority of DEGAS showed PPIs and participated in different biological processes. The top 30 DEGAS that connected with the greatest number of nodes

in the PPI network and were, therefore, considered to be the most important were identified through PPI analysis using the DMHC algorithm in Cytoscape software (Fig. 6c).

Functional enrichment analysis

GO and KEGG analyses of the 211 DEGs were performed using Metascape (Fig. 6b). The enrichment results of GO analysis performed using Metascape showed that DEGs with survival-associated AS events were mainly enriched in chordate embryonic development, response to inorganic substances, ncRNA





processing, regulation of cell–cell adhesion, neurotransmitter metabolic processes, response to oxygen levels, deoxyribonucleoside metabolic processes, nuclear division, tRNA metabolic processes, regulation of growth, cellular response to hormone stimulus, response to estradiol, carbohydrate derivative transport, cellular response to acid chemical, fatty acid metabolic processes, positive regulation of cell cycle, kinase activity, protein methyltransferase activity, regulation of cellular response to stress, and oxidoreductase activity. KEGG analysis indicated that DEGAS were mainly enriched in purine metabolism

(hsa00230), fatty acid metabolism (hsa01212), cocaine addiction (hsa05030), ovarian steroidogenesis (hsa04913), longevity regulating pathway-multiple species (hsa04213), cell cycle (hsa04110), HTLV-I infection (hsa05166), RNA degradation (hsa03018), focal adhesion (hsa04510), cytokine–cytokine receptor interaction (hsa04060), and the hippo signaling pathway (hsa04390).

Construction of prognostic models

We obtained 30 hub DEGAS from the PPI network and used them to perform a multivariate Cox

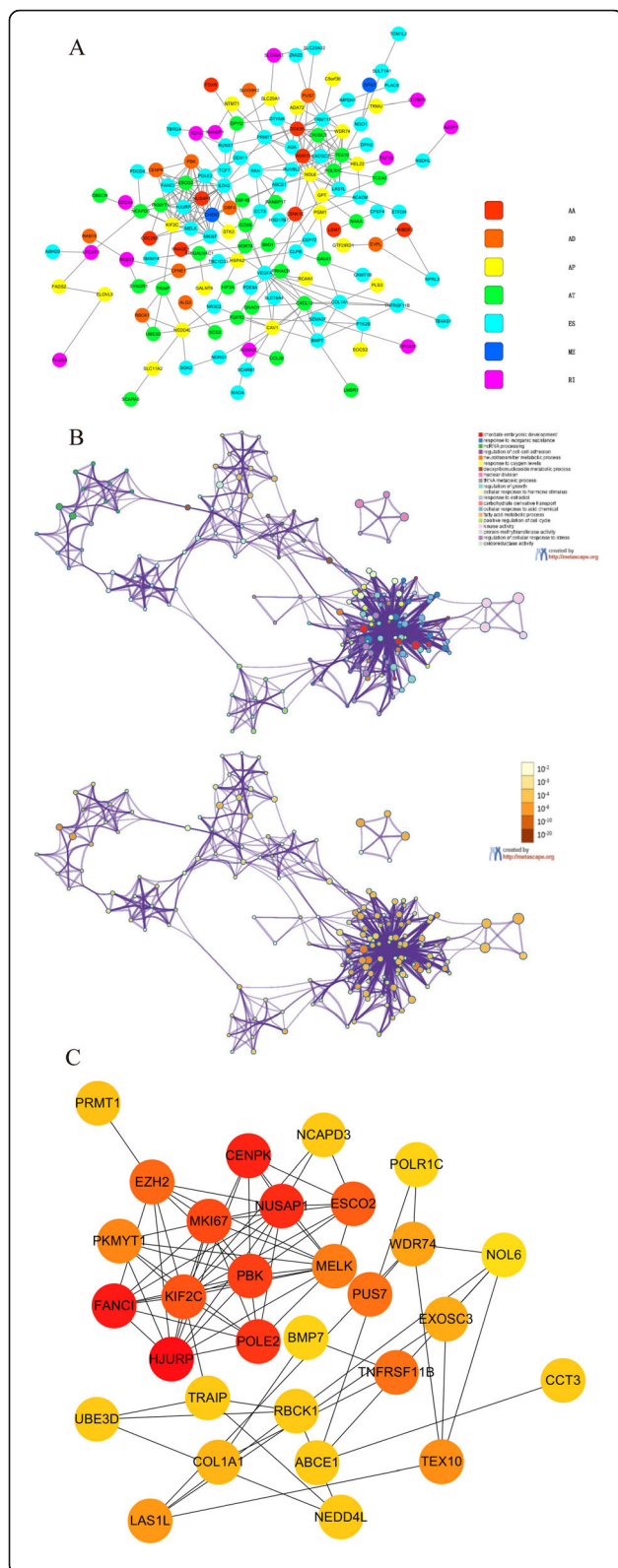


Fig. 6 Functional enrichment analysis. **a** Gene interaction networks of differentially expressed genes overlapping with alternative splicing events (DEGAS) generated by Cytoscape. **b** Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis of DEGAS generated by Metascape. **c** Top 30 most-connected hub DEGAS in the protein-protein interaction (PPI) network

regression analysis. Regression coefficients from the multivariate Cox analysis were used to construct a risk score and develop a prognostic model to predict OS. Twelve genes (*EZH2*, coefficient = - 3.36021; *NEDD4L*, coefficient = - 3.59895; *PRMT1*, coefficient = 12.4729; *LAS1L*, coefficient = 5.054636; *PBK*, coefficient = 6.15113; *KIF2C*, coefficient = - 16.7546; *NCAPD3*, coefficient = - 9.33565; *TRAIP*, coefficient = - 3.85844; *PUS7*, coefficient = - 3.51065; *COL1A1*, coefficient = 4.254234; *NUSAP1*, coefficient = - 3.00095; *MKI67*, coefficient = 1.669178) were included in the signature. General information on the 12 genes is shown in Table 1. The risk score was defined as the sum of the expression of each gene multiplied by its coefficients. The risk score was calculated for each patient. We used the median risk score as the cutoff value to classify patients into low-risk and high-risk groups. The gene expression and survival status of each patient were ranked by the risk score values of the prognostic signature (Fig. 7a). A KM curve was plotted to indicate the correlation between the risk score and OS of patients with COAD ($P_{\log\text{-rank}} = 2.188e-06$, Fig. 7b). The time-dependent ROC curve showed a promising performance in terms of OS prediction ($P < 0.05$, $AUC = 0.782$, Fig. 7c).

The results of univariate and multivariate Cox regression analysis showed that the risk score was effective at predicting the OS of patients with CRC [HR in univariate analysis = 1.051 (95% CI 1.035–1.067), HR in multivariate analysis = 1.042 (1.026–1.060); $P < 0.001$; Table 2; Fig. 7d, e]. The details of risk score together with other clinical elements based on ROC curve analysis are shown in Fig. 7g). PCA of all patients with COAD was performed to display the characteristics of each patient (Fig. 8a).

Splicing factor-regulated network construction

By analyzing level 3 mRNA-seq data from TCGA, a total of 28 splicing factors associated with the survival time of patients were identified (Fig. 8b).

Discussion

AS is a biological process that produces many different splicing isoforms from one pre-mRNA to exert various biological functions. Abnormal AS has been identified as a molecular trigger for cancer [14].

Table 1 General information of the 12 genes in the prognostic risk score

Gene stable ID	Gene name	Gene type	Chromosome	Gene start (bp)	Gene end (bp)
ENSG00000142945	KIF2C	Protein coding	1	44739704	44767767
ENSG00000183763	TRAIP	Protein coding	3	49828595	49856584
ENSG00000126457	PRMT1	Protein coding	19	49676166	49688450
ENSG00000091127	PUS7	Protein coding	7	105456501	105522271
ENSG00000001497	LAS1L	Protein coding	X	65512582	65534806
ENSG00000137804	NUSAP1	Protein coding	15	41332841	41381050
ENSG00000069399	MKI67	Protein coding	19	44742620	44760044
ENSG00000148773	PBK	Protein coding	10	128096659	128126423
ENSG00000108821	COL1A1	Protein coding	17	50184096	50201649
ENSG00000106462	EZH2	Protein coding	7	148807374	148884344
ENSG00000049759	NEDD4L	Protein coding	18	58044226	58401540
ENSG00000151503	NCAPD3	Protein coding	11	134150113	134225461

Various studies have shown that the transcript variants of a gene may have opposing roles, and AS is known to be a key factor in cancer progression, including proliferation, apoptosis, invasion, angiogenesis, and metabolism [15, 16]. For example, an EGFR splice variant, lacks exon 4 (de4 EGFR), was found to constitutively activate downstream pathways to promote cancer cell proliferation and metastasis in glioma, prostate cancer, and ovarian cancer tissues [17]. PKM2, an embryonic splice isoform resulting from mutually exclusive AS of the PKM pre-mRNA, is re-expressed in many types of tumors and regulates cancer cell metabolism and tumor growth [18]. Because of the potential significance of AS in cancer biology, its clinical relevance in malignancies has received extensive attention. Moreover, it has been clarified recently that cancer-specific splice variants may potentially be used as diagnostic, prognostic, and predictive biomarkers, as well as therapeutic targets [19]. However, previous studies have mainly focused on single genes, and there has been no comprehensive assessment of the prognostic power of AS events in COAD. In the current study, AS profiles were analyzed and used to construct a regulatory network for COAD using TCGA data. Several prognostic AS events/genes were identified, which could provide potential treatment targets for patients with COAD.

We obtained gene expression and AS profile from TCGA database which has generated large amounts of genomics and proteomics data characterizing the molecular landscape of more than 11,000 tumors across 33 cancer types. TCGA is a multi-institutional comprehensive collection of various molecularly characterized tumor types, comprises multiple levels of tumor data, including genomic, transcriptomic,

proteomic, and epigenetic data [20]. TCGA data provide unprecedented opportunities for systematic investigation of cancer mechanisms at multiple molecular and regulatory layers and has been used in numerous cancer studies [21]. Then, we investigated DEGs between normal and tumor samples. It has been shown that DEGs are related to the activation of various typical tumorigenesis-related pathways and are widely used in oncology research [22]. AS gives rise to multiple mRNA variants from a single gene, changing the expression level of the corresponding gene [23]. Therefore, we expect AS events occurring in DEGs to be particularly important in the progression of tumorigenesis, and they could, therefore, provide potential treatment targets for patients with COAD. Differential expression of mRNA amplifies or reduces the effect of AS events; therefore, we chose DEGAS as an object for research. We overlapped DEGs and survival-related AS genes to obtain DEGAS. Univariate and multivariate regression analyses were performed, and 12 DEGAS were identified. Based on 12 hub DEGAS, a prognostic risk score model was constructed. The prognostic risk score based on AS events showed a good efficacy in predicting the prognosis of patients with COAD.

Among the 12 DEGAS, *EZH2* was demonstrated to be a master regulator of cancer-associated epigenetic disorders by Makoto et al. [24]. Manipulation of *EZH2/PRC2* may, therefore, be a promising avenue for cancer therapy. Wang et al. [25] reported that *NEDD4L* plays a critical role in the regulation of cellular processes such as apoptosis, transport, and metastasis and is downregulated in several types of cancers. Song et al. [26] demonstrated that *PRMT1* expression promoted the growth of pancreatic

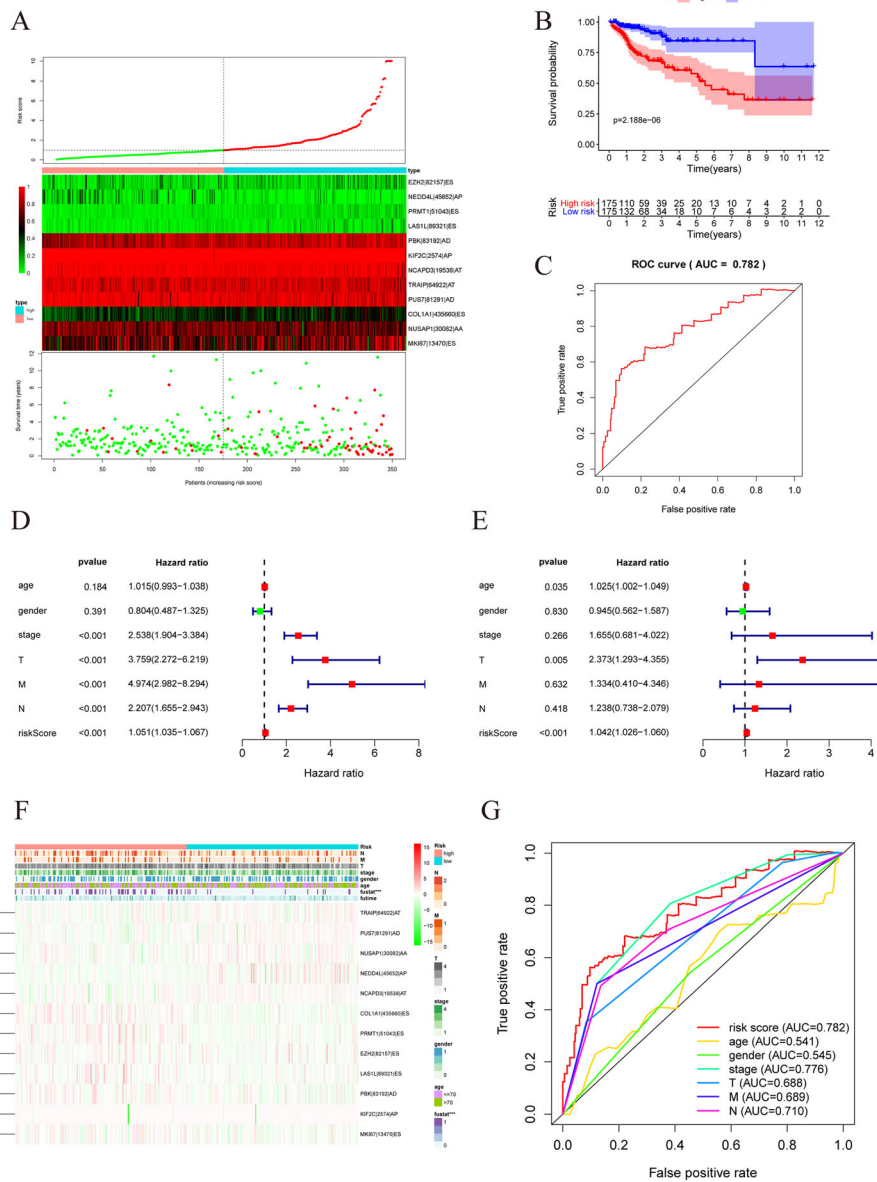


Fig. 7 Construction of prognostic risk score model. **a** Risk score plot, heatmap of prognostic differentially expressed genes overlapping with alternative splicing events (DEGAS) in risk score, survival time, and status for patients with colon adenocarcinoma (COAD). **b** Kaplan-Meier curve of risk score. **c** Receiver operating characteristic (ROC) curve of risk score. **d** Univariate Cox analysis of clinical parameters and risk scores. **e** Multivariate Cox regression analysis of clinical parameters and risk scores. **f** Clinical parameters and risk scores of patients with COAD. **g** ROC curves with area under the curve (AUC) of clinical parameters and risk score

cancer-derived cells both in vitro and in vivo and is significantly correlated with pancreatic ductal adenocarcinoma tumor size and prognosis in postoperative patients. Ma et al. demonstrated that over-expression of *PBK* promotes autophagy and enhances cisplatin resistance via the ERK/mTOR signaling pathway in high-grade serous ovarian carcinoma. Gan et al. [27] reported that *KIF2C* is a target gene of miR-325-3p, which has been shown to be a tumor suppressor in

non-small-cell lung cancer. Yuan et al. [28] reported that over-expression of *NCAPD2/3* promotes the release of pro-inflammatory cytokines by modulating the IKK/NF- κ B signaling pathway. Han et al. [29] reported that *TRAI*P is crucial in early mitotic progression and chromosome alignment defects in the metaphase. According to Brouwer et al. [30], *PUS7* is the most abundant post-transcriptional modification in RNA, which is primarily thought to stabilize

Table 2 Univariate and multivariate Cox regression analyses of the gene signature and overall survival of ccRCC patients

	N	Univariate analysis			Multivariate analysis			
		HR	95% CI	P	HR	95% CI	P	
Age	≤ 60	260	1	1.236–2.322	< 0.001	1	1.008–1.906	0.045
	> 60	247	1.694			1.384		
Gender	Female	174	1	0.694–1.318	0.787	1	0.799–1.532	0.543
	Male	333	0.957			1.106		
Stage	I/II	307	1	1.495–1.792	< 0.001	1	2.510–4.941	< 0.001
	III/IV	200	1.637			3.522		
Risk score	Low	254	1	1.060–1.118	< 0.001	1	2.017–4.202	< 0.001
	high	253	1.089			2.911		

secondary structures of RNA, and *PUS7* variants can lead to neurological defects. Liu et al. [31] showed that increased *COL1A1* expression is associated with poor survival, and *COL1A1* knockdown inhibited metastasis of breast cancer cells. Cordon et al. demonstrated that *NUSAP1* promotes invasion and metastasis of prostate cancer [32]. *MKI67*, encoding the proliferation marker protein Ki-67, is correlated with the progression of several cancers such as laryngeal squamous cell cancer, gastroenteropancreatic neuroendocrine tumors, and cervical cancer [33, 34]. These results demonstrate that most of the genes identified play an important role in the development of cancer.

In this study, we have investigated novel clinical biomarkers and therapeutic approaches, which could eventually be used to implement novel therapeutic strategies against COAD. For example, we can adjust the treatment plan by detecting the expression level of a certain molecular biomarker in the tumor site, and we can inhibit the further development of the tumor by knocking down or raising the expression level of a certain molecular biomarker. And the potential involvement of splicing factors in COAD was studied in our analysis. The splicing correlation network revealed obvious trends whereby the majority of favorable prognosis AS events were positively correlated with the expression of splicing factors, whereas adverse prognosis AS events were negatively correlated with the expression of splicing factors. Accordingly, the question of whether downregulation of specific splicing factors would give rise to reduced favorable prognosis AS events and increased adverse prognosis AS events requires further investigation through functional experiments.

Conclusions

By analyzing the relationship between AS events/genes and patient prognosis, we found that AS events may be

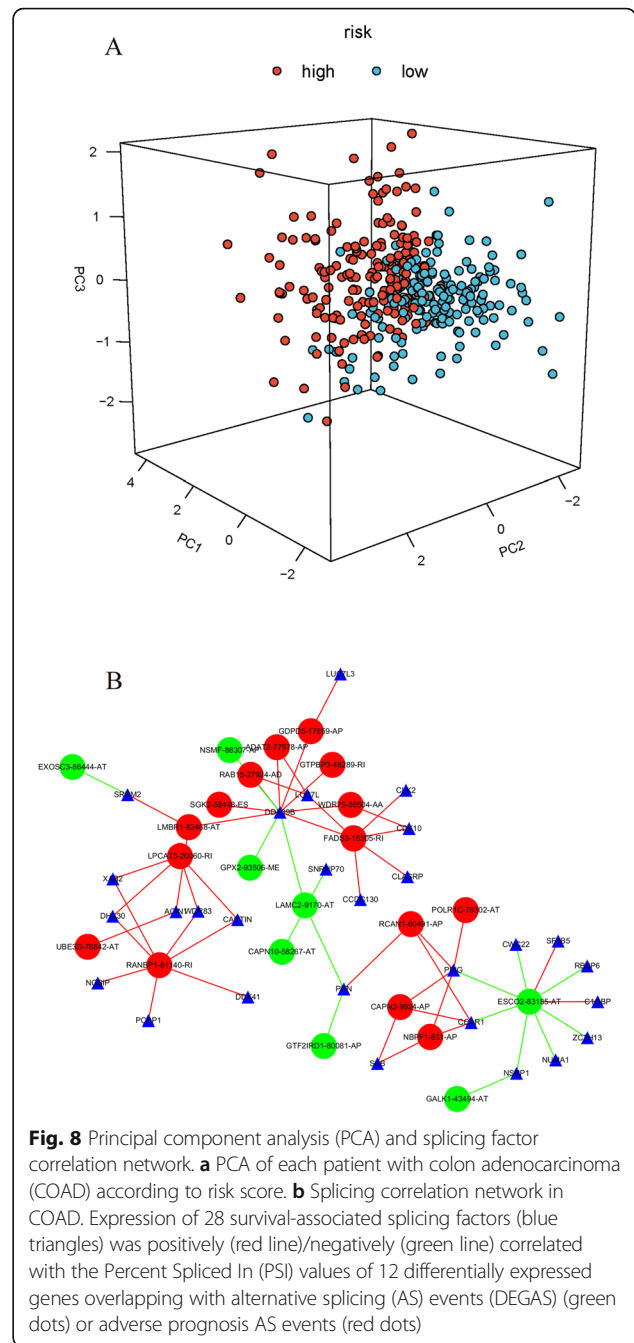


Fig. 8 Principal component analysis (PCA) and splicing factor correlation network. **a** PCA of each patient with colon adenocarcinoma (COAD) according to risk score. **b** Splicing correlation network in COAD. Expression of 28 survival-associated splicing factors (blue triangles) was positively (red line)/negatively (green line) correlated with the Percent Spliced In (PSI) values of 12 differentially expressed genes overlapping with alternative splicing (AS) events (DEGAS) (green dots) or adverse prognosis AS events (red dots)

predictors of COAD prognosis. Twelve potential genes (*EZH2*, *NEDD4L*, *PRMT1*, *LAS1L*, *PBK*, *KIF2C*, *NCAPD3*, *TRAIIP*, *PUS7*, *COL1A1*, *NUSAP1*, and *MKI67*) were identified from the interaction network and correlation analysis between AS events and gene expression. Many of these genes are involved in the development of cancer. Based on multivariate analysis for survival, the 12 genes may be used as biomarkers for the classification of COAD prognosis at AS event and gene expression levels.

Abbreviations

AA: Alternate acceptor site; AD: Alternate donor site; AP: Alternate promoter; AS: Alternative splicing; AT: Alternate terminator; AUC: Area under the curve; COAD: Colon adenocarcinoma; CRC: Colorectal cancer; DEG: Differentially expressed gene; DEGAS: Differentially expressed gene overlapping with alternative splicing event; ES: Exon skip; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; ME: Mutually exclusive exon; PPI: Protein-protein interaction; RI: Retained intron; ROC: Receiver operating characteristic; TCGA: The Cancer Genome Atlas

Acknowledgements

Not available.

Authors' contributions

(I) Conception and design: Lifei Tian and Yaqi Qu. (II) Administrative support: Lifei Tian. (III) Collection and assembly of data: Yaqi Qu, Yujia Chen. (IV) Data analysis and interpretation: Yujia Chen. (V) Manuscript writing: Yaqi Qu. (VI) Final approval of manuscript: All authors.

Funding

There is no funding.

Availability of data and materials

RNA-seq data was downloaded from the TCGA COAD cohort, and mRNA splicing pattern data was obtained by the SpliceSeq tool from TCGA. All the data was available from TCGA (<https://tcga-data.nci.nih.gov/tcga/>), and we are willing to provide our analysis data.

Ethics approval and consent to participate

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study complied with the Declaration of Helsinki and was approved by the Ethics Committees of Shaanxi Provincial People's Hospital.

Consent for publication

The authors have no interest in ethical, legal, and financial conflicts related to the article. All authors read and approved the manuscript to publish.

Competing interests

All the authors declare that they have no competing interests.

Author details

¹Department of General Surgery, Shaanxi Provincial People's Hospital, Xi'an 710068, China. ²School of Basic Medical Sciences, Hebei Medical University, Shijiazhuang, Hebei, China. ³Department of Function, Shaanxi Provincial People's Hospital, Xi'an, China.

Received: 15 May 2020 Accepted: 24 August 2020

Published online: 03 September 2020

References

- Yang Y, Zhang J, Weiss NS, et al. The consumption of chili peppers and the risk of colorectal cancer: a matched case-control study. *World J Surg Oncol*. 2019;17(1):71.
- Dong L, Ding C, Zheng T, et al. Extracellular vesicles from human umbilical cord mesenchymal stem cells treated with siRNA against ELFN1-AS1 suppress colon adenocarcinoma proliferation and migration. *Am J Transl Res*. 2019;11(11):6989–99.
- Cheng L, Eng C, Nieman LZ, Kapadia AS, Du XL. Trends in colorectal cancer incidence by anatomic site and disease stage in the United States from 1976 to 2005. *Am J Clin Oncol*. 2011;34(6):573–80.
- Chen Z, Song J, Chen L, et al. Characterization of TTN Novex Splicing Variants across Species and the Role of RBM20 in Novex-Specific Exon Splicing. *Genes (Basel)*. 2018;9(2):86.
- Wang H, Wang H, et al. Highly sensitive and multiplexed quantification of mRNA splice variants by the direct ligation of DNA probes at the exon junction and universal PCR amplification. *Chem Sci*. 2017;8(5):3635–40.
- Hsiao TH, Lin CH, Lee TT, et al. Verifying expressed transcript variants by detecting and assembling stretches of consecutive exons. *Nucleic Acids Res*. 2010;38(20):e187.
- Zhang C, Dower K, Zhang B, et al. Computational identification and validation of alternative splicing in ZSF1 rat RNA-seq data, a preclinical model for type 2 diabetic nephropathy. *Sci Rep*. 2018;8(1):7624.
- Sakuma K, Sasaki E, Kimura K, et al. HNRNPLL stabilizes mRNA for DNA replication proteins and promotes cell cycle progression in colorectal cancer cells. *Cancer Sci*. 2018;109(8):2458–68.
- Fischer PA, Baas H. Therapeutic efficacy of R(-)-deprenyl as adjuvant therapy in advanced parkinsonism. *J Neural Transm Suppl*. 1987;25:137–47.
- Guo Y, Bao Y, Yang W. Regulatory miRNAs in colorectal carcinogenesis and metastasis. *Int J Mol Sci*. 2017;18(4):890.
- Yoo W, Shah DV, Chih MY, Gustafson DH. A smartphone-based support group for alcoholism: effects of giving and receiving emotional support on coping self-efficacy and risky drinking. *Health Informatics J*. 2020;26(3):1764–76. PMID: 31814490.
- Liu X, Liu L, Dong Z, et al. Expression patterns and prognostic value of m(6)A-related genes in colorectal cancer. *Am J Transl Res*. 2019;11(7):3972–91.
- Chen X, Wang C, Jiang Y, et al. Bcl-3 promotes Wnt signaling by maintaining the acetylation of β -catenin at lysine 49 in colorectal cancer. *Signal Transduct Target Ther*. 2020;5(1):52.
- Lee CC, Chang WH, Chang YS, et al. 4 β -Hydroxywithanolide E modulates alternative splicing of apoptotic genes in human hepatocellular carcinoma Huh-7 cells. *Sci Rep*. 2017;7(1):7290.
- Genov N, Basti A, Abreu M, Relógio A. Temporal splicing switches in elements of the TNF-pathway identified by computational analysis of transcriptome data for human cell lines. *Int J Mol Sci*. 2019;20(5):1182.
- Siveen KS, Prabhu KS, Paray AS, et al. Evaluation of cationic channel TRPV2 as a novel biomarker and therapeutic target in Leukemia-Implications concerning the resolution of pulmonary inflammation. *Sci Rep*. 2019;9(1):1554.
- Wang H, Zhou M, Shi B, et al. Identification of an exon 4-deletion variant of epidermal growth factor receptor with increased metastasis-promoting capacity. *Neoplasia*. 2011;13(5):461–71.
- Yang W, Lu Z. Regulation and function of pyruvate kinase M2 in cancer. *Cancer Lett*. 2013;339(2):153–8.
- Cherry S, Lynch KW. Alternative splicing and cancer: insights, opportunities, and challenges from an expanding view of the transcriptome. *Genes Dev*. 2020;34(15-16):1005–16.
- Piergiorgio F, Claudia C, Claudio B, et al. Genetic research: the role of citizens, public health and international stakeholders. *TOPHJ-12-106* 2019.
- Cannovo N, Guarino R, Fedeli P. Ethical and deontological aspects of pediatric biobanks: the situation in Italy. *Cell Tissue Bank*. 2020;21(3):469–77.
- Pizzollo J, Nielsen WJ, Shibata Y, et al. Comparative serum challenges show divergent patterns of gene expression and open chromatin in human and chimpanzee. *Genome Biol Evol*. 2018;10(3):826–39.
- Hernando CE, García Hourquet M, de Leone MJ, et al. A role for Pre-mRNA-processing protein 40C in the control of growth, development, and stress tolerance in *Arabidopsis thaliana*. *Front Plant Sci*. 2019;10:1019.
- Yamagishi M, Uchimarum K. Targeting EZH2 in cancer therapy. *Curr Opin Oncol*. 2017;29(5):375–81.
- Wang X, Duan J, Fu W, et al. Decreased expression of NEDD4L contributes to NSCLC progression and metastasis. *Biochem Biophys Res Commun*. 2019; 513(2):398–404.
- Song C, Chen T, He L, et al. PRMT1 promotes pancreatic cancer growth and predicts poor prognosis. *Cell Oncol (Dordr)*. 2020;43(1):51–62. PMID: 31520395.
- Gan H, Lin L, Hu N, et al. KIF2C exerts an oncogenic role in nonsmall cell lung cancer and is negatively regulated by miR-325-3p. *Cell Biochem Funct*. 2019;37(6):424–31.
- Yuan CW, Sun XL, Qiao LC, et al. Non-SMC condensin I complex subunit D2 and non-SMC condensin II complex subunit D3 induces inflammation via the IKK/NF- κ B pathway in ulcerative colitis. *World J Gastroenterol*. 2019; 25(47):6813–22.
- Park IS, Jo KS, Won HS, Kim H. Dimerization of TRAF-interacting protein (TRAIIP) regulates the mitotic progression. *Biochem Biophys Res Commun*. 2015;463(4):864–9.
- de Brouwer APM, Abou JR, Kortel N, et al. Variants in PUS7 cause intellectual disability with speech delay, microcephaly, short stature, and aggressive behavior. *Am J Hum Genet*. 2018;103(6):1045–52.
- Liu J, Shen JX, Wu HT, et al. Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target. *Discov Med*. 2018; 25(139):211–23.

32. Gordon CA, Gong X, Ganesh D, Brooks JD. NUSAP1 promotes invasion and metastasis of prostate cancer. *Oncotarget*. 2017;8(18):29935–50.
33. Nowinska K, Chmielewska M, Piotrowska A, et al. Correlation between levels of expression of minichromosome maintenance proteins, Ki-67 proliferation antigen and metallothionein I/II in laryngeal squamous cell cancer. *Int J Oncol*. 2016;48(2):635–45.
34. Krtnic D, Zivadinovic R, Jovic Z, et al. Significance of the Ki-67 proliferation index in the assessment of the therapeutic response to cisplatin-based chemotherapy in patients with advanced cervical cancer. *Eur Rev Med Pharmacol Sci*. 2018;22(16):5149–55.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

