

Comprehensive profiling of aberrant prognostic alternative splicing signatures in clear cell renal cell carcinoma: a retrospective study based on large-scale sequencing data

Dong Zhang

Shandong University Qilu Hospital

Zhao Zhang

Shandong University Qilu Hospital

Yi Duan

Shandong University Qilu Hospital

Guangxu Ji

Junan County People's Hospital

Hongliang Wu

Shandong University Qilu Hospital

Yueqing Tang

Shandong University Qilu Hospital

Xiaolin Hu

Jinan Central Hospital Affiliated to Shandong University

Zhiqing Fang

Shandong University Qilu Hospital

Zhaoxin Guo (✉ sdlzx2k@126.com)

Qilu Hospital of Shandong University <https://orcid.org/0000-0002-1655-3457>

Hainan Liu

Shandong University Qilu Hospital

Zhaoxu Liu

Shandong University Qilu Hospital

Zhonghua Xu

Shandong University Qilu Hospital

Primary research

Keywords: ccRCC, alternative splicing, TCGA, prognosis, splicing factors

Posted Date: June 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-33113/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Clear cell renal cell carcinoma(ccRCC) is the most common type with poor prognosis in kidney tumor. Growing evidence has indicated that aberrant alternative splicing (AS) events are efficacious signatures for tumor prognosis predicting and therapeutic targets. Systematic and comprehensive analysis of AS in ccRCC is in urgent need.

Methods: Level 3 RNA-seq data were acquired from TCGA data portal and the AS profiles were performed with assistance of SpliceSeq software. Univariate cox regression analysis was applied for screening prognosis-related AS events. Gene functional enrichment analysis revealed the pathways enriched by prognosis-related AS. The final AS panel was developed by LASSO-penalized method for predicting prognosis and compared with traditional clinical factors. The potential regulatory network was analyzed via Spearman correlation between splicing factors (SFs).

Results: A total of 2100 survival-associated AS events were filtered from 1666 parent genes. Gene functional enrichment analysis suggested that the regulation of autophagy could be a potential mechanism of splicing regulatory in ccRCC. 17 aberrant AS events formed the final AS panel which can estimate OS probability in ccRCC patients. The AUC values of ROC curves for the final AS panel can keep above 0.7 spanning 1 year to 5 years.

Conclusion: We developed a robust and individualized predictive model based on large-scale sequencing data. The identified vital AS events and splicing networks may be valuable in deciphering the potential mechanisms of AS on tumorigenesis of ccRCC.

Background

Renal cell carcinoma (RCC), originated from the renal epithelium, is the most common form of renal cancer and accounts for >90% of cancers in the kidney. Annually, over 290,000 new cases are diagnosed and approximate 134,000 deaths are recorded worldwide[1]. The disease encompasses several subtypes, of which clear cell RCC (ccRCC) is the most prevalent and contributes to most cancer-related deaths[2]. Partial or radical nephrectomy is one of the effective methods to treat ccRCC. However, 20–30% of patients treated with surgery will relapse, despite having no evidence of metastases when diagnosed with ccRCC[3; 4]. Similar to other solid tumors, ccRCC development and progression are characterized by aberrant genetic and protein expression. In the past decade, genetic alterations of ccRCC have been extensively investigated, revealing that specific genetic and epigenetic alterations are associated with carcinogenesis, recurrence and metastasis of tumor[1; 5; 6]. Hence, it is crucial to explore molecular mechanisms more deeply and excavate vital biomarkers which can improve prognosis of ccRCC.

Protein diversity is fundamental for generating remarkable regulatory and functional complexity of human cells. A basic mechanism for protein diversity is the alternative modification and processing of pre-mRNA[7]. Alternative splicing (AS) is a post-transcriptional process which occurs in >90% human genes, engendering an enormous number of new isoforms from few sets of genes by inclusion or

exclusion of different exons or parts of exons in pre-mRNA. This process has a significant impact on the diversity of both transcriptome and proteome of human[8; 9]. AS plays a vital role in increasing the complexity of functional proteins and regulating of cell metabolism and cell-specific functions, which provides opportunity for tumorigenesis and progression[10]. Over the last two decades, molecular tools have been developed to correct or redirect AS events. It has been observed that a switch on particular AS event could occur in cancer related genes[11]. Aberrant AS can lead to loss-of-function in tumor suppressors or activation of oncogenes[12]. Moreover, researches show that splicing factors can not only regulate AS events directly, but also can be regulated by various signaling pathways, making AS sensitive to tumor microenvironment[13; 14]. Accordingly, deciphering valuable AS events may help to elucidate underlying mechanisms of oncogenesis in ccRCC and provide new therapeutic approaches.

Most current researches of AS in tumor have focused on finding cancer-specific AS events by distinguishing tumor tissues from normal controls[15; 16]. As far as we know, the prognostic value of abnormal AS events in ccRCC has not been comprehensively studied. In our study, we systematically analyzed genome-wide AS events in ccRCC cohort by mining The Cancer Genome Atlas (TCGA) database and further constructed a powerful predictive model based on a series of survival-related AS events. This will provide a novel insight on exploring the pathogenesis of ccRCC and make a significant impact on personalized precision-based care in patients.

Materials And Methods

2.1. Data acquisition and pre-processing

Level 3 RNA-seq data and the corresponding clinical datasheets for ccRCC samples were downloaded from GDC data portal of the Cancer Genome Atlas (TCGA) website (<https://portal.gdc.cancer.gov/>)[17; 18]. Analysis of AS profiles was performed with assistance of SpliceSeq software and TCGA SpliceSeq database, which evaluated the mRNA splicing pattern and quantified AS events with the Percent Spliced In (PSI) score calculated, ranging between 0 and 1[19; 20]. To generate a reliable set of AS events, we implemented the following filter criteria: percentage of samples with PSI score ≥ 75 , PSI range across samples ≥ 0.05 and standard deviation of PSI score ≥ 0.02 . And the remaining AS events with not available (Null) were imputed via the k-Nearest Neighbor (kNN) imputation algorithm[21]. Besides, our subsequent analyses were based on ccRCC cohort comprised of patients that meet the inclusion criteria: (1) a histological diagnosis of ccRCC; (2) patients who did not receive prior treatment, especially neoadjuvant chemotherapy; (3) patients with relatively complete clinical features including gender, age, tumor site, Furhman grade, pathological stage, T stage, M stage and overall survival (OS) information; (4) the follow-up time was no less than 30 days; (5) patients with corresponding mRNA splicing profiles. The present study fully satisfies the TCGA publication guidelines.

2.2. Differentially spliced AS events analysis

To identify differentially expressed AS events within each AS type between ccRCC and normal tissues, student's t-test, followed by an adjustment of p-value using the Benjamini-Hochberg correlation for

multiple comparison, was applied. The $|\log_2FC| > 1$ and adjusted p-value < 0.05 indicated the AS events were significant differences between tumor and adjacent normal tissues.

2.3. Identification of survival-associated AS events and functional enrichment analyses

After rigorous screening, a total of 455 ccRCC patients with differentially expressed AS profiles and survival information were subjected to subsequent analyses. For each specific AS event, patients were divided into two group based on median cut. A univariate Cox proportional hazard regression analysis was used to define the prognostic value of each AS event. In this analysis, AS events were regarded as significant at p-value < 0.05 . And Upset plot, generated by UpsetR, was visualized to quantitatively analyze the gene interactions among the seven types of prognostic AS events[22]. Then to further shed light on the potential modifying mechanism of aberrant AS events on corresponding protein, the parent genes of these survival-associated AS events were sent for functional enrichment analyses via clusterProfiler R package, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis[23]. The GO terms and KEGG pathways were considered significant with a threshold of false discovery rate (FDR) < 0.05 .

2.4. Dimension reduction and generation of AS signature

For avoiding multicollinearity of highly correlated variables, the Cox regression model, with least absolute shrinkage and selection operator (LASSO) penalty, was implemented to reduce dimension[24]. Firstly, among the top 20 most significant AS events in univariate analysis within seven AS types, the key prognostic biomarkers were selected to further develop prognostic signatures via LASSO-penalized method. Then we utilized the regression coefficients derived from fitting multivariate Cox model to multiply the PSI scores of prognostic indicators for constructing AS signature in each AS type. Furthermore, the candidate AS events from seven signatures were combined together to build the final AS panel for ccRCC cohort. Then we divided patients into low-risk and high-risk subgroups based on median value of each model, the Kaplan-Meier survival analysis and log-rank test were applied to verify the prognostic ability of prognostic models. Moreover, the predictive efficiency of each model was assessed by time-dependent receiver operator characteristic (ROC) curves. Therein, the LASSO Cox model was constructed with the “glmnet” R package, using the 10-fold cross-validation and “lambda.min” criteria such that the tuning parameters (lambda) were at the minimum[25]. The area under curve (AUC) of time-dependent ROC curve was calculated with “timeROC” R package[26].

2.5. Independence of final AS panel from clinicopathological features

To investigate the independent prognostic value of final AS panel from conventional clinicopathological characteristics (including age, gender, Furhman grade, pathological stage, T stage, M stage) in ccRCC patients, univariate followed by multivariate Cox regression analyses were conducted. Then to further confirm whether the final AS panel was of high applicability and robust in various subgroups, stratification Cox analyses were also performed.

2.6. Construction and evaluation of the nomogram

In order to provide surgeons with a clinically relevant quantitative approach for predicting the short-term and long-term survival probability of ccRCC patients individually, we assembled a nomogram that integrated the final AS panel and independent clinicopathological risk factors based on the above results of multivariate analysis. Then the calibration curves were graphically assessed to determine whether the derived nomogram performed well compared to ideal model. Therein, the nomogram and calibration plot were depicted using rms package [27; 28]. And Harrell's concordance index (C-index) was calculated to estimate the discrimination ability of the nomogram with the help of Hmisc package. Moreover, the clinical usefulness of the nomogram, the AS panel and clinical risk factors were compared with the aid of decision curve analysis (DCA) [29; 30; 31; 32].

2.7. Gene set enrichment analysis (GSEA) of final AS panel

To further identify crucial biological processes and cancer-specific pathways related to the final AS panel, we performed a GSEA using the adjusted expression profiles for all transcripts. The annotated gene set files of "c2.cp.v6.2.symbols" and "c5.bp.v6.2.symbols" download from the "Molecular Signatures Databases (<http://software.broadinstitute.org/gsea/msigdb>)" were employed for running GSEA using the java software (<http://software.broadinstitute.org/gsea/downloads.jsp>) [33; 34]. Enrichment p-values were based on 1000 permutations and the significance threshold was set at nominal p-value < 0.05.

2.8. Construction of splicing regulatory network

A list of 88 human splicing factors (SFs) was created by integrating SpliceAid 2 database (www.introni.it/spliceaid.html) and the work from Xiong et al, which collected the experimentally validated SFs through hand-curated screening of literature and databases [35; 36]. First, we compared ccRCC samples and adjacent normal samples to identify differentially expressed SFs using student's t-test. Then we assessed the correlation between the normalized expression value (variance stabilizing transformation via DESeq2 package) of SFs and OS through fitting univariate regression analysis in the entire cohort, where the SFs with p-value < 0.05 were selected as prognostic ones for further analysis [37]. The "surv_cutpoint" function of the "survminer" R package was used to iteratively determine the optimal cutpoints of prognostic SFs achieving the maximally selected rank statistics. Then Spearman correlation analyses were performed between expression value of prognostic SFs and PSI scores of the most significant AS events in each AS type. P-values were adjusted by Benjamini-Hochberg (BH) procedure and the significance threshold was set at an adjusted p-value < 0.05.

Results

3.1. Overview of integrated AS events profiling in TCGA ccRCC cohort

After a series of stringent filtering and screening, a total of 455 ccRCC patients from the TCGA with integrated mRNA splicing variants profiling and counterpart clinical information were retrospectively

analyzed in depth. The included cohort was comprised of 155 female (34.1%) and 300 male (65.9%) patients, among which 145 (31.9%) died and the median survival time was 79.5 months. The median follow-up time of these patients with ccRCC was 37.0 months (range, 1.3-112.6 months). For the entire cohort, the 1-, 3-, 5-year OS rates were 89.9%, 74.7% and 61.0%, respectively. Besides, we detected and quantifies the AS events of transcripts by using SpliceSeq software. According to the exclusive and distinct splicing pattern, these AS events can be roughly divided into seven particular types, including Exon Skip (ES), Alternate Terminator (AT), Alternate Promoter (AP), Retained Intron (RI), Mutually Exclusive Exons (ME), Alternate Donor site (AD) and Alternate Acceptor site (AA). As results, we obtained 32487 mRNA splicing events from 9286 parent genes, which contains 2634 AAs in 1964 genes, 2366 ADs in 1751 genes, 6155 ATs in 2806 genes, 173 MEs in 167 genes, 11371 ESs in 5230 genes, 2031 RIs in 1404 genes and 7757 APs in 3259 genes. And ES (35%), as the predominant type, accounts for more than one-third of all events, followed by AP and AT type in ccRCC and paracancerous tissues. We also noticed that one gene might possess two or more events. Considering this, Upset plot was generated to quantitatively visualize the intersecting sets among these seven AS types (Supplementary Fig. S1). Intriguingly, the vast majority of splicing variants were from the same genes while one gene could have up to five types of AS events, revealing the fact that AS hold largest potential on expanding transcript diversity and increasing proteome complexity.

3.2. Identification of aberrant cancer-specific AS events in ccRCC

Considering the significantly spliced AS events between tumor and adjacent normal samples could play crucial roles in the tumorigenesis and development of ccRCC, we performed differential expression analyses with 72 normal tissues and 533 TCGA ccRCC tissues summarized. Using the preliminarily screening criteria mentioned above, we derived a list of 4660 aberrant splicing events, including 1069 ESs in 823 genes, 543 RIs in 438 genes, 1777 APs in 1470 genes, 719 ATs in 676 genes, 231 ADs in 204 genes, 304 AAs in 280 genes and 17 MEs in 17 genes. Interestingly, ES only contributes about 22.9% aberrant events even though it constitutes the highest proportion of AS events. As is shown in Fig. 1, most differential expressed AS events tend to be up-regulated in ccRCC tissues, suggesting the dysregulated splicing variants prefer to generate oncogenic transcripts rather than silence tumor suppressor transcripts.

3.3. Construction of AS panel and evaluation of its predictive ability in ccRCC cohort

Moreover, we attempted to assess the prognostic ability of these identified aberrant AS events. By using univariate Cox regression analyses, we identified a total of 2100 survival-associated AS events from 1666 parent genes (p -value < 0.05), including 1975 unfavorable prognostic splicing events (hazard ratio (HR) > 1, p -value < 0.05) and 125 favorable prognostic ones (HR < 1, p -value < 0.05). The exact number of aberrant events indicating poor prognosis far exceeded the number of those indicating good prognosis, which were consistent with the results from differential analysis and appear to confirm a major potential oncogenic role of aberrant AS in initiation and progression of ccRCC. To further quantify the interactions among these aberrant survival-associated events, we applied Upset plot and revealed that some genes

clearly had up to five types of AS events which were both cancer-specific and prognosis-related (Fig. 2A). For example, AD, AP, ES, RI and AA events of DMKN (red dotted line) were all significantly associated with patient OS in ccRCC. NDRG2 and POFUT2 (green dotted lines) could possess four types of prognostic AS events. In addition, we presented the top 20 most significantly survival-associated events in each AS type (only 5 ME events available) in Fig. 2B-H, respectively.

Furthermore, these top most significant ones were chosen as candidates for predictive model construction. To discover the specific AS events with the greatest prognostic value, we applied LASSO-penalized regression model with 10-fold cross validation to the candidate AS events in seven types, separately (Supplementary Figs. S2 and S3). In our data analysis, based on OS information, we have generated seven AS signatures, which were comprised of 4 AA events, 6 AD events, 10 AP events, 7 AT events, 11 ES events, 4 RI events and 3 ME events, respectively. Then patients were stratified into low-risk and high-risk subgroups based on the corresponding risk score, utilizing the median cutoff calculated in the entire cohort, as demonstrated in Supplementary Fig. S4. The distribution of patients' survival time and status, heatmaps of selected splicing events were also displayed. The Kaplan-Meier survival curves also indicated that patients of high-risk were with significantly poor OS compared than those of low-risk (Supplementary Fig. S5, p-value < 0.0001). Eventually, with the combination of all available types of AS events that formed signatures, we screened 17 particular AS events (see detailed information in Supplementary Figure S6) via LASSO-penalized Cox method (Fig. 3A and Fig. 3B) and further constructed a final AS-based panel by fitting multivariate Cox regression model (Fig. 3C). In addition, the median risk score also served as the cutoff point for assigning ccRCC patients into low- and high-risk groups. As shown in Fig. 3D, the high-risk patients showed a 3.13-fold higher risk (95% confidence intervals (CI): 2.18-4.50, p-value < 0.0001) than low-risk patients. Encouragingly, time-dependent ROC curves (Fig. 4A-C) also revealed that the AUC values of established AS panel were 0.73 at 1 year, 0.724 at 3 years and 0.75 at 5 years, confirming the predictive potential of our AS panel for both short-term and long-term prognosis with AUCs being robust above 0.7 across time (Fig. 4D). The detailed formula of each AS signature was listed in Supplementary Table S1.

3.4. Functional enrichment results of identified aberrant survival-associated AS events

It was evident that AS has a profound effect in altering transcript architecture and further modifying corresponding protein via deleting or adding functional domains. Thus, analyzing the influencing of aberrant AS on their encoding proteins might give us a better understanding and direction of the roles of aberrant events in the splicing machinery of ccRCC. In our research, 1666 genes from 2100 both ccRCC-specific and OS-related splicing events were sent for bioinformatics analysis, including GO and KEGG pathway enrichment analyses. As shown in Supplementary Table S2, a total of 48 GO-CC categories, 60 GO-BP and 25 GO-MF terms were identified as significantly enriched ones (FDR < 0.05), most of which were closely related to mitochondrial and energy metabolic activities, such as mitochondrial matrix (FDR < 0.0001), generation of precursor metabolites and energy (FDR = 0.003), and mitochondrial inner membrane (FDR = 0.002). Fig. 5 also shows that specific biological processes AS closely associated with ccRCC, such as focal adhesion (FDR = 0.002), RNA splicing (FDR = 0.015), and GTPase regulatory

activities, were significantly enriched in these parent genes occurred aberrant AS events. Besides, several KEGG pathways (Fig. 5D) that were experimentally validated participating in ccRCC progression or metastasis, including central carbon metabolism in cancer (FDR = 0.011), ubiquitin mediated proteolysis (FDR = 0.011), base excision repair (FDR = 0.043) and apoptosis (FDR = 0.016), were also revealed. Of note, both KEGG and GO enrichment analyses suggested the regulation of autophagy could be a potential mechanism of splicing regulatory in ccRCC. Taken together, above functional enrichment results not only provided novel clues for further exploring the underlying mechanisms upon aberrant prognosis-related events, but also confirmed the reliability of our screening procedure in turn.

3.5. The independent predictive power of final AS panel for ccRCC patients

We performed univariate followed by multivariate Cox hazard regression analyses of data in TCGA ccRCC cohort in order to further investigate whether the final 17 AS based panel was an independent prognostic factor, where the AS panel was treated as a binary variable. The univariate analysis results suggested that age, Fuhrman grade, pathological stage, T stage, M stage and final AS panel were all remarkably correlated with the OS of ccRCC patients (Fig. 6A). Therefore, those significant factor were included in a multivariate analysis, which showed that the AS panel (HR = 2.487; 95% CI: 1.717-3.605; p-value = 1.47e-06), age, Fuhrman grade, pathological stage were four independent prognostic factors when adjusted by those risk factors (Fig. 6B).

Moreover, in order to investigate the prognostic value of our AS panel in stratified cohort, we classified patients into various subgroups based on age, gender, pathological stage, Fuhrman grade, T stage and M stage then we performed stratification analysis. As shown in Fig. 7, the AS panel identified patients with distinct prognoses in all cohorts analyzed, thus confirming its robustness for independently predicting ccRCC prognosis.

3.6. Altered biological pathways between the high- and low-risk ccRCC patients

The strong risk stratification ability of our 17 AS-based panel for ccRCC patients could be attributed to the oncogenic pathways and biological process that participate in ccRCC regulation during the tumor development or metastasis. Hence, we performed GSEA to elucidate potential biological functions of the final AS panel (Fig. 3E and Fig. 3F). In the GSEA enrichment results, we observed that genes highly expressed in the high-risk group showed significantly enriched in "P53 signaling transduction" and "MAPK signaling pathway", which have been widely acknowledged being associated with the progression of ccRCC. Furthermore, several cancer-related pathways, such as the cell cycle regulation, response to DNA damage and neutrophil chemotaxis, were also up-regulated in high-risk patients, while the high-risk correlated genes were revealed to be associated with down-regulation of autophagy, which was in accordance with the KEGG enrichment results above. In summary, our GSEA analysis implied that the final 17 AS-based signature might be involved in crucial ccRCC-related pathways and their functional dysregulation subtly influences the OS of ccRCC patients.

3.7. Development and apparent performance of a nomogram based on the final AS panel

A second multivariate Cox regression model that incorporated the above independent predictors, including the final AS panel, age, Furhman grade and pathological stage, was developed and presented as the nomogram (Fig. 8A). It substantiated that the AS panel comprised of 17 AS events contributes the most risk points (ranging from 0 to 100), whereas the other clinicopathological characteristics contribute much less. The C-index for OS prediction was 0.816 with 1000 bootstrap replicates (95% CI: 0.784-0.848), suggesting derived nomogram showed good predictive discrimination ability. And the calibration plot showed that the bias-corrected lines were close to the ideal line (45° line), indicating an optimal agreement between the survival prediction by our nomogram and actual observation (Fig. 8B). Moreover, we conducted DCA analysis to determine the real-world clinical usefulness of the AS-clinicopathological nomogram by quantifying the net benefits against a range of threshold probabilities in ccRCC cohort. And the DCA results are presented in Fig. 8C, which showed that using our AS-based nomogram would add more benefits in prognosis prediction than either treat-all-patients or treat-none scheme if the threshold probability of a patients or clinician is more than 10%. Within this range, the net benefits were comparable, with several overlaps, based on the final AS panel and the model with clinicopathological factors integrated only, further convincing the AS panel is non-inferior to clinical features. Besides, the clinical usefulness of the nomogram significantly overwhelmed the clinical risk factors. In a word, these findings suggested that the derived nomogram was a better prognostic model for predicting short-term and long-term survival probability in ccRCC patients than individual clinical indicators.

3.8. The prognostic SFs and construction of splicing regulatory network

It has been proved that the global dysregulation of AS events could be orchestrated by a limited number of SFs, especially in ccRCC. The normalized level 3 RNA-seq profiles of TCGA ccRCC samples identified 53 SFs whose expression levels differed significantly between tumor tissues and adjacent normal tissues (Fig. 9A). According to the thresholds we set, 14 SFs, including 5 risky ones and 9 SFs indicating favorable prognosis, meeting the cancer-specific and prognosis-related properties were further screened as candidates (Fig. 9B and Supplementary Fig. S7). Next, to systematically analyze the cancer-specific splicing regulatory connections between SFs and AS events in ccRCC, a splicing regulatory network (Fig. 9C), which enrolled the significant relationships (adjusted.p.value < 0.05) among these 14 SFs (blue and yellow dots) and 110 most significantly prognostic AS events comprised of 105 unfavorable ones (red dots) and 5 favorable ones (green dots), was finally built. From the regulatory network, we can see that the majority of genes could be synergistically or competitively regulated by different SFs via binding the same transcript region (AS event), further revealing the molecular complexity and tumor heterogeneity in ccRCC. Indeed, the entire SF-AS regulatory network was quite intricate, and altered SFs played vital driving roles in modifying pathological splicing events. Representative scatter plots presenting the significant correlations between expression levels of particular SFs and PSI scores of specific AS events were displayed in Fig. 9D. For instance, the expression of ESRP2 was positively correlated with AD event of ZFP36, while the ES of SLC25A29 could be negatively regulated by SRP54.

Discussion

As changeable and possibly heritable post-transcriptional process, AS mechanisms offer promising clues for the therapeutic strategies in various disease, including many cancers[11]. To date, accumulated evidence indicates the AS process could make crucial contribution to the initiation, progression and metastasis of cancer, especially in renal cell carcinoma. For instance, DCLK1-ASVs, as an alternate splice variant of DCLK1, plays a stem cell supportive role in kidney cancer via driving self-renewal and drug resistance to chemotherapy[38]. VEGF is widely regarded the most potent growth factor for tumor neovasculature, while in-depth study has revealed that VEGF_{165b}, a differential splicing isoform from exon 7 to 3' untranslated region, is down-regulated in renal cell carcinoma and contributes to anti-angiogenesis[39]. Similarly, PKM2, rather than PKM1, has been reported to promote different aspects of ccRCC progression, including cell proliferation, invasion and migration in vitro[40]. Meanwhile, the involvement of SFs is also reflected in the investigation of aberrant AS regulation in ccRCC. The pro-tumorigenic function of SF3B3 can be partly mediated by up-regulation of EZH2Δ14, since the EZH2Δ14 acts as a suppressor for the expression of EZH2, which is commonly recognized as a tumor-suppressive gene[41]. Accordingly, AS events have been considered as multifaceted carcinogenesis hallmark of ccRCC, convincing that splicing events could be preminent biomarkers for further investigation. Even so, painstaking exploration of therapies concerning aberrant AS, including clinical trials, are nevertheless under way[14; 42]. To our best knowledge, our present study is a relatively comprehensive attempt to decipher the untapped mechanisms of AS in ccRCC so far.

The prevalence of ccRCC, accounting for approximately 75% of primary kidney cancer, is increasing annually with recent implementation of screening program. Despite the advance of surgical procedures, such as partial or total nephrectomy, more than 20% postoperative patients suffering from ccRCC will develop distant metastasis[3; 43]. What's worse, adjuvant systematic therapies, including chemotherapy, radiotherapy and immunotherapy, have failed to show an obvious survival benefit for postoperative patients, highlighting the drug resistance and financial burden as problems worth pondering[1; 44]. In this aspect, novel molecular biomarkers that could reliably evaluate disease progression and patient prognosis would have tremendous potential in guiding therapeutic strategies and clinical management of ccRCC patients. Taking the potential influence of aberrant splicing events in renal carcinogenesis into account, it is reasonable to consider the AS events in to increase the effectiveness of adjuvant therapy. In this study, using the well-established public, large-scale multi-omics sequencing cohort, we proposed a robust, individualized AS panel that can estimate OS probability in ccRCC patients based on 17 aberrant splicing events. Attributed to the combination of all the seven types of AS events, the AUC values of ROC curves for this model can keep above 0.7 spanning 1 year to 5 years, suggesting enhanced power and great potential in short-term and long-term prognosis prediction for ccRCC patients. Moreover, our prognostic AS panel could further stratify clinically defined groups of patients (e.g. age, TNM stage and Fuhrman grade) into subgroups with distinct prognostic outcomes.

Actually, to illustrate the clinical value of AS events in cancer, several investigators have performed genome-wide analyses of splicing events in non-small cell lung cancer, gastrointestinal adenocarcinoma and ovarian cancer[45; 46; 47]. With the rapid development of high-throughput sequencing technologies

over the last decades, preliminary success has also been gained in insight into the involvement of splicing patterns in ccRCC. Recently, by integrating whole-exome sequencing, RNA-seq and copy number variation analyses, Lehmann et al analyzed 282 ccRCC patients and revealed that 915 probable splicing quantitative trait loci with tumor-specific splicing pattern were identified, and that are involved in processes relevant for tumor growth and cancer progression[48]. However, few studies have broken the gap between aberrant AS events and clinical meaning in a systematic way. And it still remains to be unsolved problem that what extent alteration of AS pattern could change the clinical outcome among ccRCC patients. Despite a series of strict criteria were conducted in the process of sample screening and biomarker filtration, more AS events, consisting of a total of 2100 aberrant splicing events occurred in 1666 parent genes, were finally identified in present study, meeting the cancer-specific and prognosis-related properties. Moreover, our exploration of AS patterns broadens the understanding of traditionally conceptual prognostic biomarker, which serves as a powerful complementary to verify the link between AS and ccRCC.

Furthermore, the functional enrichment analyses, including GO, KEGG and GSEA analysis, provided more opportunities for deciphering the largely untapped mechanisms which our identified AS events might participate in. Intriguingly, there were several intersections among our enrichment results and preliminary investigation. The majority of enriched biological behaviors were metabolism-related and the autophagy-related pathways were revealed as the significant overlap among GO, KEGG and GSEA results. Autophagy, as an intracellular self-degradative process for capturing and recycling degraded components with the assistance of autophagosome and lysosome, plays crucial roles in maintaining metabolism and homeostasis[49]. More importantly, the modulation of autophagy involves in dual-sided roles within cancer microenvironment, which could facilitate tumorigenesis via promoting proliferation and mediate tumor suppression via inducing cell death and apoptosis[50]. ccRCC is distinguished by inactivating mutations in VHL (von Hippel-Lindau tumor suppressor), leading to constitutive activation of the hypoxia-inducible factors (HIFs) and induction of a hypoxia response transcription signature. Hall et al found TRPM3 and miR-204 established a regulatory circuit which could promote the growth of ccRCC and stimulate MAP1LC3A (LC3A) and MAP1LC3B (LC3B) autophagy. VHL represses TRPM3 directly through miR-204 and indirectly through another miR-204 target[51].

SF-AS regulatory network may provide new approaches to address the underlying mechanism of the splicing pathway involved in ccRCC. Here, we are particularly interested in ESRP2 and hnRNPs. ESRP2 is a kind of splicing regulator specifically expressed in epithelial cells, regulating of the function of cytoskeleton and cell motility. Mizutani et al reported that the expression of some ESRP-target exons was related with good prognosis in ccRCC and with the expression of RNF111, a molecular which interacted with ESRP2, induced polyubiquitination and suppressed ccRCC tumor growth in a coordinated manner[52]. hnRNPs regulate almost all levels of expression of apoptotic genes and play a vital role in DNA repair, telomere biogenesis, cell signaling, mRNA stability, alternative splicing, and protein degradation. Through these key cellular functions, individual hnRNPs have a variety of potential roles in tumor development and progression including the inhibition of apoptosis, angiogenesis and cell invasion[53].

With the advance of the times and the increasing demand for precision medicine, the limitations of traditional predictive factors have gradually emerged. They may be insufficient for proper tumor characterization and classification, leading to difficulties for the prediction of disease progression, therapeutic response and prognosis. Nowadays, powerful and advanced technologies including high-throughput, genome-wide profiling methodologies has led to significant progress in the biological understanding and prognosis of ccRCC[54]. Here, we demonstrated that our AS panel could serve as an independent prognostic factor for OS prediction in ccRCC patients after the modification of traditional risk factors. Besides, we further leveraged the complementary value of AS panel and other independent clinic pathological features (age, Furhrman grade and pathologic stage). Consequently, our nomogram could be a valuable novel prognostic scoring system for clinical decision-making of practitioners.

Even though current comprehensive profiling of aberrant prognostic AS signatures provided us an AS-clinicopathologic nomogram, which represented as the robust and reliable tool for prognosis prediction in ccRCC, several limitations were still required to be clarified. First, a major limitation of our work is due to its retrospective nature, and clearly, multicentre prospective studies with large sample size are warranted to further validate our findings, especially the established AS panel. Second, owing to the existence of missing value for clinicopathological characteristics and a fraction of splicing events, the statistical power in integrated prognostic model could be inevitably decreased. Third, our AS panel is developed by numerous AS events, where further biological experiments are needed to verify their functions in ccRCC.

Conclusion

In summary, for the first time, we identified an AS panel composed of 17 aberrant AS events in ccRCC at the genome-wide level. The interpretation of aberrant AS patterns and corresponding regulatory networks may reveal a new underlying mechanism in the oncogenic process of ccRCC and provide novel molecular alterations and potential drug targets that require further validation.

Declarations

Acknowledgements

Not applicable.

Authors' Contributions

Zhaoxin Guo and Dong Zhang designed the study. Dong Zhang performed the data analysis. Zhao Zhang and Yi Duan wrote the article. Guangxu Ji and Hongliang Wu designed the figures and tables. Xiaolin Hu and Yueqing Tang carried out the primary literature search. Hainan Liu proofread the article format. Zhaoxin Guo, Zhiqing Fang, and Zhonghua Xu revised and edited the article. All authors read and approved the final article.

Funding Information

This study was supported by the Key Research Project of Shandong Province (No. 2017GSF18162).

Availability of data and materials

All data in the current study are based on public data available in The Cancer Genome Atlas (TCGA) datasets.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have approved the submitted manuscript.

Competing interests

The authors report no conflicts of interest in this work.

References

- [1] J.J. Hsieh, M.P. Purdue, S. Signoretti, C. Swanton, L. Albiges, M. Schmidinger, D.Y. Heng, J. Larkin, and V. Ficarra, Renal cell carcinoma. *Nat Rev Dis Primers* 3 (2017) 17009.
- [2] B.A. Inman, M.R. Harrison, and D.J. George, Novel immunotherapeutic strategies in development for renal cell carcinoma. *Eur Urol* 63 (2013) 881-9.
- [3] M. Meskawi, M. Sun, Q.D. Trinh, M. Bianchi, J. Hansen, Z. Tian, M. Rink, S. Ismail, S.F. Shariat, F. Montorsi, P. Perrotte, and P.I. Karakiewicz, A review of integrated staging systems for renal cell carcinoma. *Eur Urol* 62 (2012) 303-14.
- [4] I. Wolff, M. May, B. Hoschke, R. Zigeuner, L. Cindolo, G. Hutterer, L. Schips, O. De Cobelli, B. Rocco, C. De Nunzio, A. Tubaro, I. Coman, B. Feciche, M. Truss, O. Dalpiaz, R.S. Figenshau, K. Madison, M. Sanchez-Chapado, M.D. Santiago Martin, L. Salzano, G. Lotrecchiano, S.F. Shariat, M. Hohenfellner, R. Waidelich, C. Stief, K. Miller, S. Pahernik, S. Brookman-May, C.P. Members of the, and G. the European Association of Urology Young Academic Urologists Renal Cancer, Do we need new high-risk criteria for surgically treated renal cancer patients to improve the outcome of future clinical trials in the adjuvant setting? Results of a comprehensive analysis based on the multicenter CORONA database. *Eur J Surg Oncol* 42 (2016) 744-50.
- [5] J.J. Hsieh, V. Le, D. Cao, E.H. Cheng, and C.J. Creighton, Genomic classifications of renal cell carcinoma: a critical step towards the future application of personalized kidney cancer care with pan-omics precision. *J Pathol* 244 (2018) 525-537.
- [6] C.J. Ricketts, A.A. De Cubas, H. Fan, C.C. Smith, M. Lang, E. Reznik, R. Bowlby, E.A. Gibb, R. Akbani, R. Beroukhim, D.P. Bottaro, T.K. Choueiri, R.A. Gibbs, A.K. Godwin, S. Haake, A.A. Hakimi, E.P. Henske, J.J.

Hsieh, T.H. Ho, R.S. Kanchi, B. Krishnan, D.J. Kwiatkowski, W. Lui, M.J. Merino, G.B. Mills, J. Myers, M.L. Nickerson, V.E. Reuter, L.S. Schmidt, C.S. Shelley, H. Shen, B. Shuch, S. Signoretti, R. Srinivasan, P. Tamboli, G. Thomas, B.G. Vincent, C.D. Vocke, D.A. Wheeler, L. Yang, W.Y. Kim, A.G. Robertson, N. Cancer Genome Atlas Research, P.T. Spellman, W.K. Rathmell, and W.M. Linehan, The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* 23 (2018) 313-326 e5.

[7] T.W. Nilsen, and B.R. Graveley, Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463 (2010) 457-63.

[8] F.E. Baralle, and J. Giudice, Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* 18 (2017) 437-451.

[9] Y. Wang, J. Liu, B.O. Huang, Y.M. Xu, J. Li, L.F. Huang, J. Lin, J. Zhang, Q.H. Min, W.M. Yang, and X.Z. Wang, Mechanism of alternative splicing and its regulation. *Biomed Rep* 3 (2015) 152-158.

[10] L.M. Gallego-Paez, M.C. Bordone, A.C. Leote, N. Saraiva-Agostinho, M. Ascensao-Ferreira, and N.L. Barbosa-Morais, Alternative splicing: the pledge, the turn, and the prestige : The key role of alternative splicing in human biological systems. *Hum Genet* 136 (2017) 1015-1042.

[11] N. Martinez-Montiel, N.H. Rosas-Murrieta, M. Anaya Ruiz, E. Monjaraz-Guzman, and R. Martinez-Contreras, Alternative Splicing as a Target for Cancer Treatment. *Int J Mol Sci* 19 (2018).

[12] J. Chen, and W.A. Weiss, Alternative splicing in cancer: implications for biology and therapy. *Oncogene* 34 (2015) 1-14.

[13] A.R. Kornblihtt, I.E. Schor, M. Allo, G. Dujardin, E. Petrillo, and M.J. Munoz, Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* 14 (2013) 153-65.

[14] X. Song, Z. Zeng, H. Wei, and Z. Wang, Alternative splicing in cancers: From aberrant regulation to new therapeutics. *Semin Cell Dev Biol* 75 (2018) 13-22.

[15] Y.S. Tsai, D. Dominguez, S.M. Gomez, and Z. Wang, Transcriptome-wide identification and study of cancer-specific splicing events across multiple tumors. *Oncotarget* 6 (2015) 6825-39.

[16] E. Sebestyen, M. Zawisza, and E. Eyra, Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res* 43 (2015) 1345-56.

[17] K. Tomczak, P. Czerwinska, and M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* 19 (2015) A68-77.

[18] A. Colaprico, T.C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T.S. Sabedot, T.M. Malta, S.M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, and H. Noushmehr, TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 44 (2016) e71.

- [19] M.C. Ryan, J. Cleland, R. Kim, W.C. Wong, and J.N. Weinstein, SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics* 28 (2012) 2385-7.
- [20] M. Ryan, W.C. Wong, R. Brown, R. Akbani, X. Su, B. Broom, J. Melott, and J. Weinstein, TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res* 44 (2016) D1018-22.
- [21] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R.B.J.B. Altman, Missing value estimation methods for DNA microarrays. *Bioinformatics* 17 (2001) 520-525.
- [22] J.R. Conway, A. Lex, and N. Gehlenborg, UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33 (2017) 2938-2940.
- [23] G. Yu, L.G. Wang, Y. Han, and Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (2012) 284-7.
- [24] J.H. Friedman, T. Hastie, and R.J.J.o.S.S. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent. *Biometrika* 97 (2010) 149-167.
- [25] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw* 39 (2011) 1-13.
- [26] P. Blanche, J.F. Dartigues, and H. Jacqmin-Gadda, Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 32 (2013) 5381-97.
- [27] V.P. Balachandran, M. Gonen, J.J. Smith, and R.P. DeMatteo, Nomograms in oncology: more than meets the eye. *The Lancet Oncology* 16 (2015) e173-e180.
- [28] A. Iasonos, D. Schrag, G.V. Raj, and K.S. Panageas, How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 26 (2008) 1364-70.
- [29] E.W. Steyerberg, and A.J. Vickers, Decision curve analysis: a discussion. *Med Decis Making* 28 (2008) 146-9.
- [30] A.J. Vickers, and E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 26 (2006) 565-74.
- [31] B. Van Calster, L. Wynants, J.F.M. Verbeek, J.Y. Verbakel, E. Christodoulou, A.J. Vickers, M.J. Roobol, and E.W. Steyerberg, Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol* 74 (2018) 796-804.
- [32] M. Fitzgerald, B.R. Saville, and R.J.J.J.o.t.A.M.A. Lewis, Decision Curve Analysis. *Stat Med* 34 (2015) 409-10.

- [33] S. Aravind, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *J Proceedings of the National Academy of Sciences of the United States of America*. 43 (2005).
- [34] A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, J.P. Mesirov, and P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 1 (2015) 417-425.
- [35] F. Piva, M. Giulietti, A.B. Burini, and G. Principato, SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum Mutat* 33 (2012) 81-5.
- [36] Y. Xiong, Y. Deng, K. Wang, H. Zhou, X. Zheng, L. Si, and Z. Fu, Profiles of alternative splicing in colorectal cancer and their clinical significance: A study based on large-scale sequencing data. *EBioMedicine* 36 (2018) 183-195.
- [37] M.I. Love, W. Huber, and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15 (2014) 550.
- [38] Y. Ge, N. Weygant, D. Qu, R. May, W.L. Berry, J. Yao, P. Chandrakesan, W. Zheng, L. Zhao, K.L. Zhao, M. Drake, K.J. Vega, M.S. Bronze, J.J. Tomasek, G. An, and C.W. Houchen, Alternative splice variants of DCLK1 mark cancer stem cells, promote self-renewal and drug-resistance, and can be targeted to inhibit tumorigenesis in kidney cancer. *Int J Cancer* 143 (2018) 1162-1175.
- [39] D.J.C.R. Bates, VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, is down-regulated in renal cell carcinoma. 62 (2002).
- [40] J. Jiang, X. Chen, H. Liu, J. Shao, R. Xie, P. Gu, and C. Duan, Polypyrimidine Tract-Binding Protein 1 promotes proliferation, migration and invasion in clear-cell renal cell carcinoma by regulating alternative splicing of PKM. 7 (2017) 245.
- [41] K. Chen, H. Xiao, J. Zeng, G. Yu, H. Zhou, C. Huang, W. Yao, W. Xiao, J. Hu, W. Guan, L. Wu, J. Huang, Q. Huang, H. Xu, and Z. Ye, Alternative Splicing of EZH2 pre-mRNA by SF3B3 Contributes to the Tumorigenic Potential of Renal Cancer. *Clin Cancer Res* 23 (2017) 3428-3441.
- [42] C. Bianchi, S. Bombelli, F. Raimondo, B. Torsello, V. Angeloni, S. Ferrero, V. Di Stefano, C. Chinello, I. Cifola, L. Invernizzi, P. Brambilla, F. Magni, M. Pitto, G. Zanetti, P. Mocarelli, and R.A. Perego, Primary cell cultures from human renal cortex and renal-cell carcinoma evidence a differential expression of two spliced isoforms of Annexin A3. *Am J Pathol* 176 (2010) 1660-70.
- [43] X. Feng, L. Zhang, W. Tu, and S. Cang, Frequency, incidence and survival outcomes of clear cell renal cell carcinoma in the United States from 1973 to 2014: A SEER-based analysis. *Medicine (Baltimore)* 98 (2019) e16684.
- [44] M.B. Atkins, and N.M. Tannir, Current and emerging therapies for first-line treatment of metastatic clear cell renal cell carcinoma. *Cancer Treat Rev* 70 (2018) 127-137.

- [45] M. Qu, J. Yu, H. Liu, Y. Ren, C. Ma, X. Bu, and Q. Lan, The Candidate Tumor Suppressor Gene SLC8A2 Inhibits Invasion, Angiogenesis and Growth of Glioblastoma. *Mol Cells* 40 (2017) 761-772.
- [46] P. Lin, R.Q. He, F.C. Ma, L. Liang, Y. He, H. Yang, Y.W. Dang, and G. Chen, Systematic Analysis of Survival-Associated Alternative Splicing Signatures in Gastrointestinal Pan-Adenocarcinomas. *EBioMedicine* 34 (2018) 46-60.
- [47] J. Zhu, Z. Chen, and L. Yong, Systematic profiling of alternative splicing signature reveals prognostic predictor for ovarian cancer. *Gynecol Oncol* 148 (2018) 368-374.
- [48] K.V. Lehmann, A. Kahles, C. Kandath, W. Lee, and G.J.P.S.o.B.P.S.o.B. Räscht, Integrative genome-wide analysis of the determinants of RNA splicing in kidney renal clear cell carcinoma. 20 (2014) 44-55.
- [49] C.W. Yun, and S.H. Lee, The Roles of Autophagy in Cancer. *Int J Mol Sci* 19 (2018).
- [50] P. Rajendran, A.M. Alzahrani, H.N. Hanieh, S.A. Kumar, R. Ben Ammar, T. Rengarajan, and M.A. Alhoot, Autophagy and senescence: A new insight in selected human diseases. *J Cell Physiol* 234 (2019) 21485-21492.
- [51] D.P. Hall, N.G. Cost, S. Hegde, E. Kellner, O. Mikhaylova, Y. Stratton, B. Ehmer, W.A. Abplanalp, R. Pandey, J. Biesiada, C. Harteneck, D.R. Plas, J. Meller, and M.F. Czyzyk-Krzeska, TRPM3 and miR-204 establish a regulatory circuit that controls oncogenic autophagy in clear cell renal cell carcinoma. *Cancer Cell* 26 (2014) 738-53.
- [52] A. Mizutani, D. Koinuma, H. Seimiya, and K. Miyazono, The Arkadia-ESRP2 axis suppresses tumor progression: analyses in clear-cell renal cell carcinoma. *Oncogene* 35 (2016) 3514-23.
- [53] B. Carpenter, C. MacKay, A. Alnabulsi, M. MacKay, C. Telfer, W.T. Melvin, and G.I. Murray, The roles of heterogeneous nuclear ribonucleoproteins in tumour development and progression. *Biochim Biophys Acta* 1765 (2006) 85-100.
- [54] M. Rydzanicz, T. Wrzesinski, H.A. Bluyssen, and J. Wesoly, Genomics and epigenomics of clear cell renal cell carcinoma: recent developments and potential applications. *Cancer Lett* 341 (2013) 111-26.

Figures

that are differentially expressed and survival-associated in ccRCC. (b-h) Forest plot which displays the top 20 (if available) most significantly survival-associated AS events in each type, respectively. P-values of univariate Cox regression analyses are indicated by the color scale by the side. Unadjusted hazard ratios (boxes) and 95% confidence intervals (horizontal lines) are depicted in the plot.

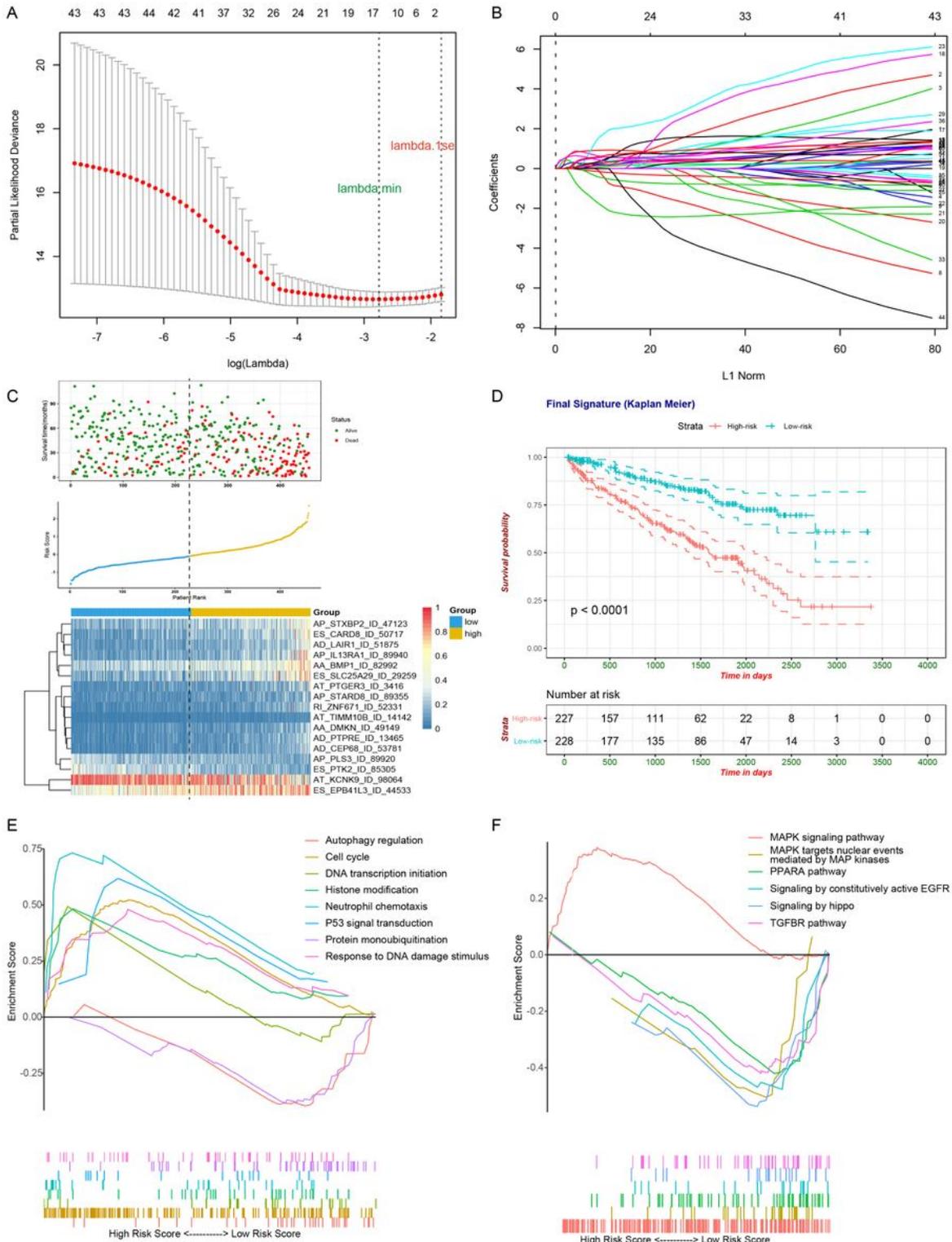


Figure 3

. Development of final AS signature by combining all types of AS events. (a) The selection of tuning parameter (λ) in the LASSO-penalized Cox regression model via 10-fold cross validation for constructing final prognostic signature. (b) The shrinkage procedure of LASSO regression coefficients to determine the optimal number of features in final model. (c) The distribution of patient survival status ranked by corresponding risk score, the splicing pattern of specific AS events included in final signature. (d) The Kaplan-Meier survival curves of overall survival between high-risk and low-risk patients, which are divided by the median value of final risk score. (e) Gene set enrichment analysis (GSEA) of hallmark gene sets named "c5.bp.v6.2.symbols" downloaded from Molecular Signatures Databases. (f) GSEA analysis delineates the biological processes related to final signature using the "c2.cp.v6.2.symbols" gene sets downloaded from Molecular Signatures Databases.

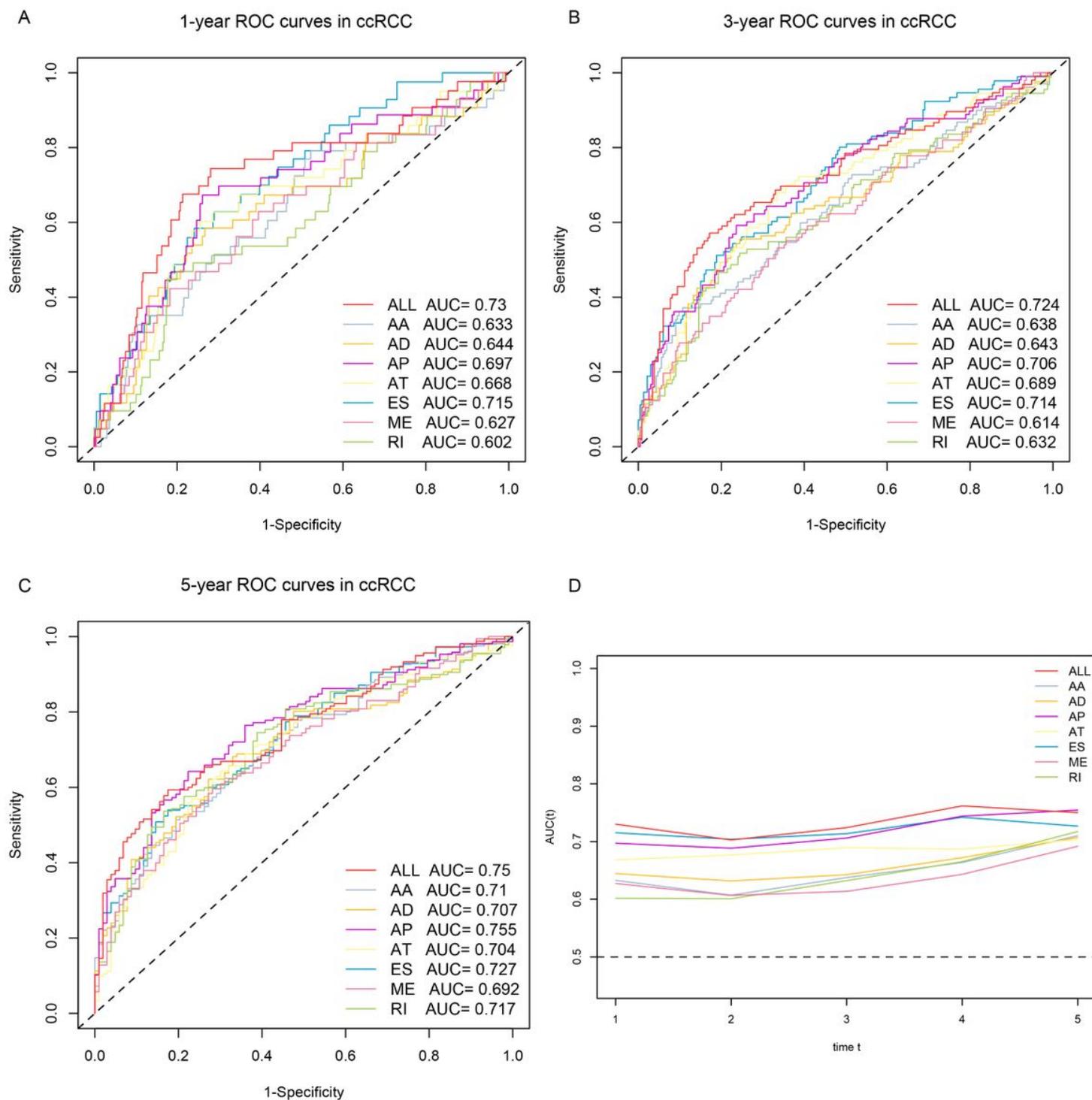


Figure 4

. The performance of the models in ccRCC cohort, including the seven AS signature and final prognostic signature. (a-c) Time-dependent ROC curves of each AS signature and final signature for predicting overall survival of ccRCC patients in 1, 3 and 5 years. (d) The comparable AUC values spanning 1 year to 5 years according to each signature.

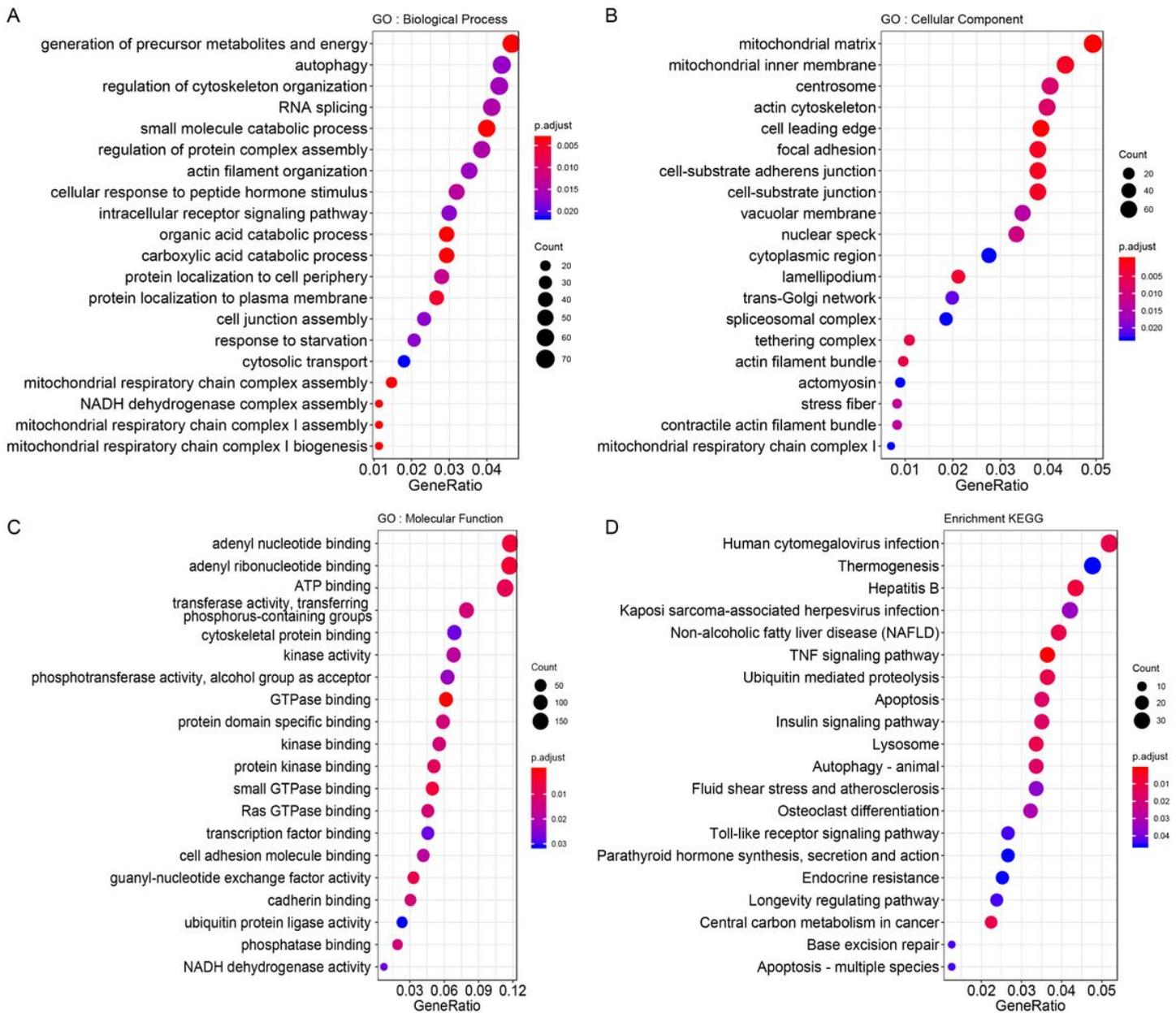


Figure 5

The functional enrichment analyses of parent genes which occurred aberrant survival-associated AS events in ccRCC patients. (a-c) Bubble diagrams showing the top significant terms from Gene Ontology (GO) analyses of corresponding genes occurred aberrant survival-associated AS events, at aspect of (a) biological process, (b) cellular component and (c) molecular function. (d) The most significant pathways identified with Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.

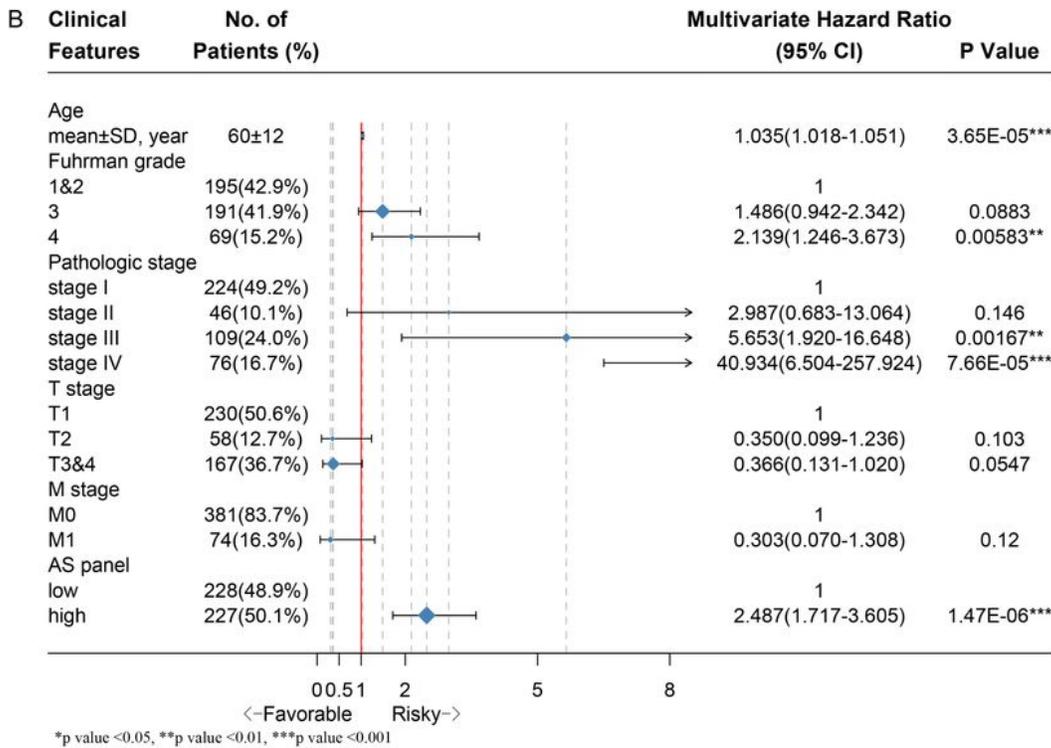
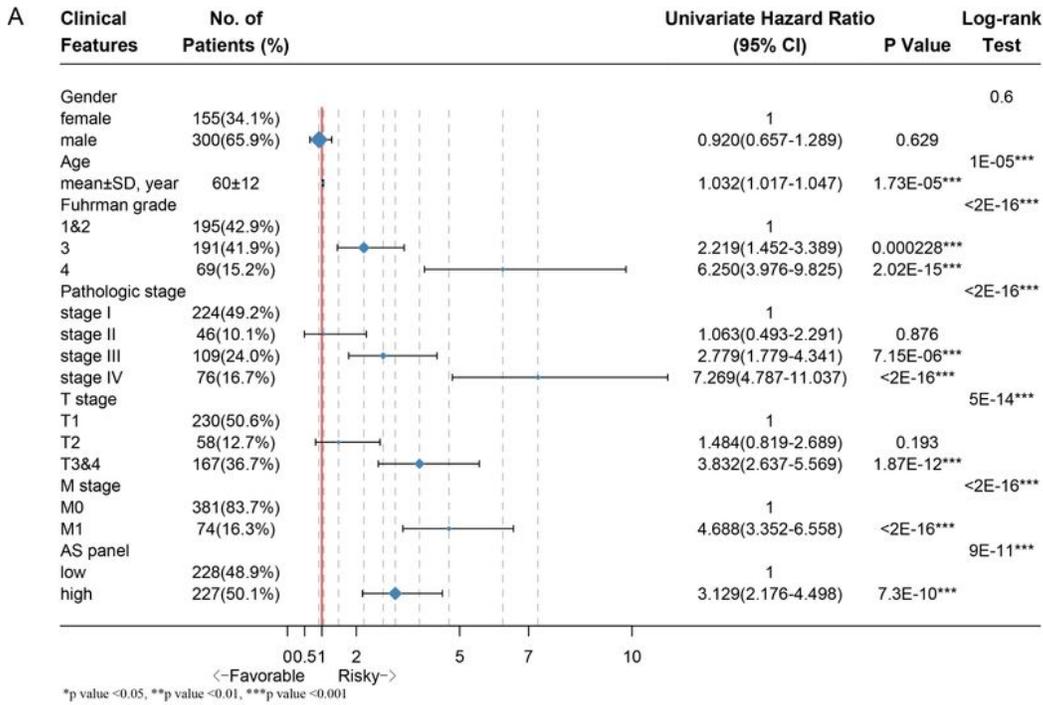


Figure 6

Forest plots of the univariate and multivariate Cox regression analysis in ccRCC cohort. (a) Univariate Cox analysis of available clinicopathologic features and final AS signature on overall survival. (b) Multivariate regression analysis of the relation between final AS prognostic model and clinicopathologic characteristics regarding prognostic value. The length of horizontal line corresponds to the 95%

confidence interval (CI), and the relative size of box is proportional to the sample size. The red vertical line indicates HR of 1.0. (* p-value < 0.05, ** p-value < 0.01 and *** p-value < 0.001).

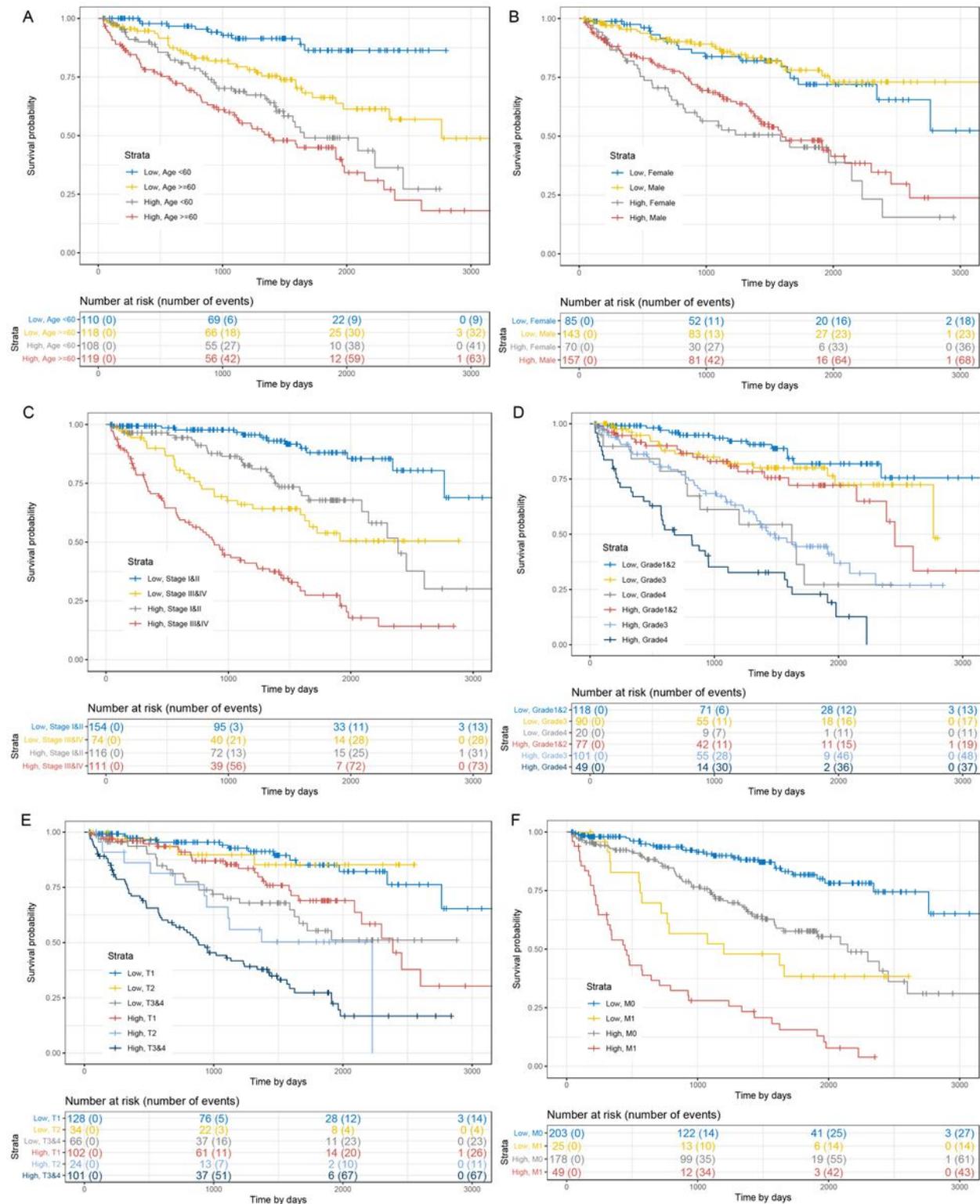


Figure 7

Stratification analyses of final AS signature in ccRCC cohort. Kaplan-Meier curves of overall survival for ccRCC patients in subgroup stratified by (a) age, (b) gender, (c) pathologic stage, (d) Fuhrman grade, (e) T

stage, (f) M stage, showing that the survival of patients was significantly poorer in patients with high risk in stratified cohorts.

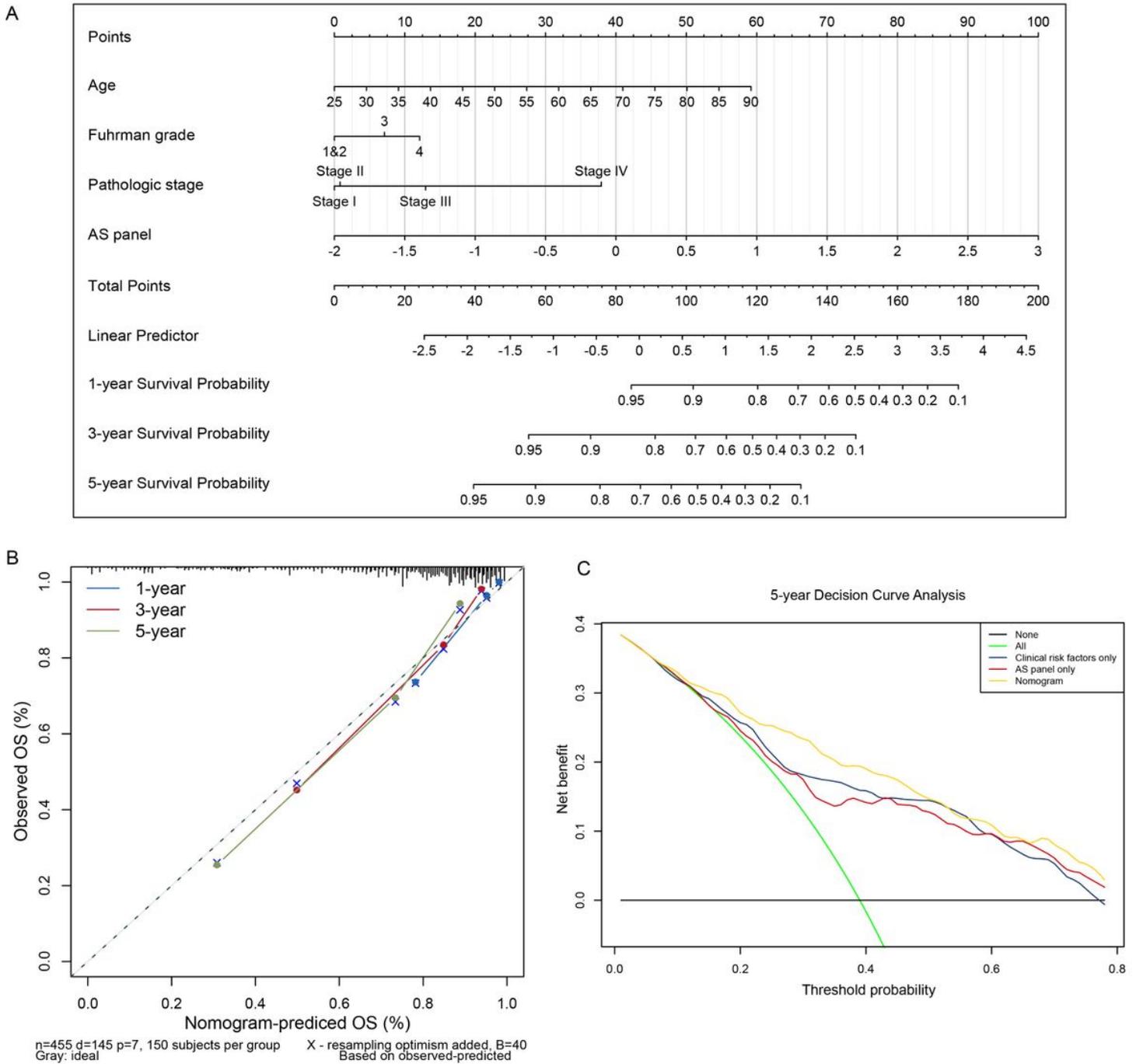


Figure 8

Construction and evaluation of a clinical prediction model. (a) A nomogram for predicting the survival probability of 1-, 3- and 5-year overall survival for ccRCC patients, which was built based on four independent prognostic factors in ccRCC. (b) Plot depicts the calibration of the nomogram based on overall survival in terms of agreement between predicted and observed 1-, 3- and 5-year clinical

outcomes. The dashed 45-degree line indicates the ideal prediction. (c) Decision curve analysis for evaluating 5-year clinical usefulness of nomogram in ccRCC patients.

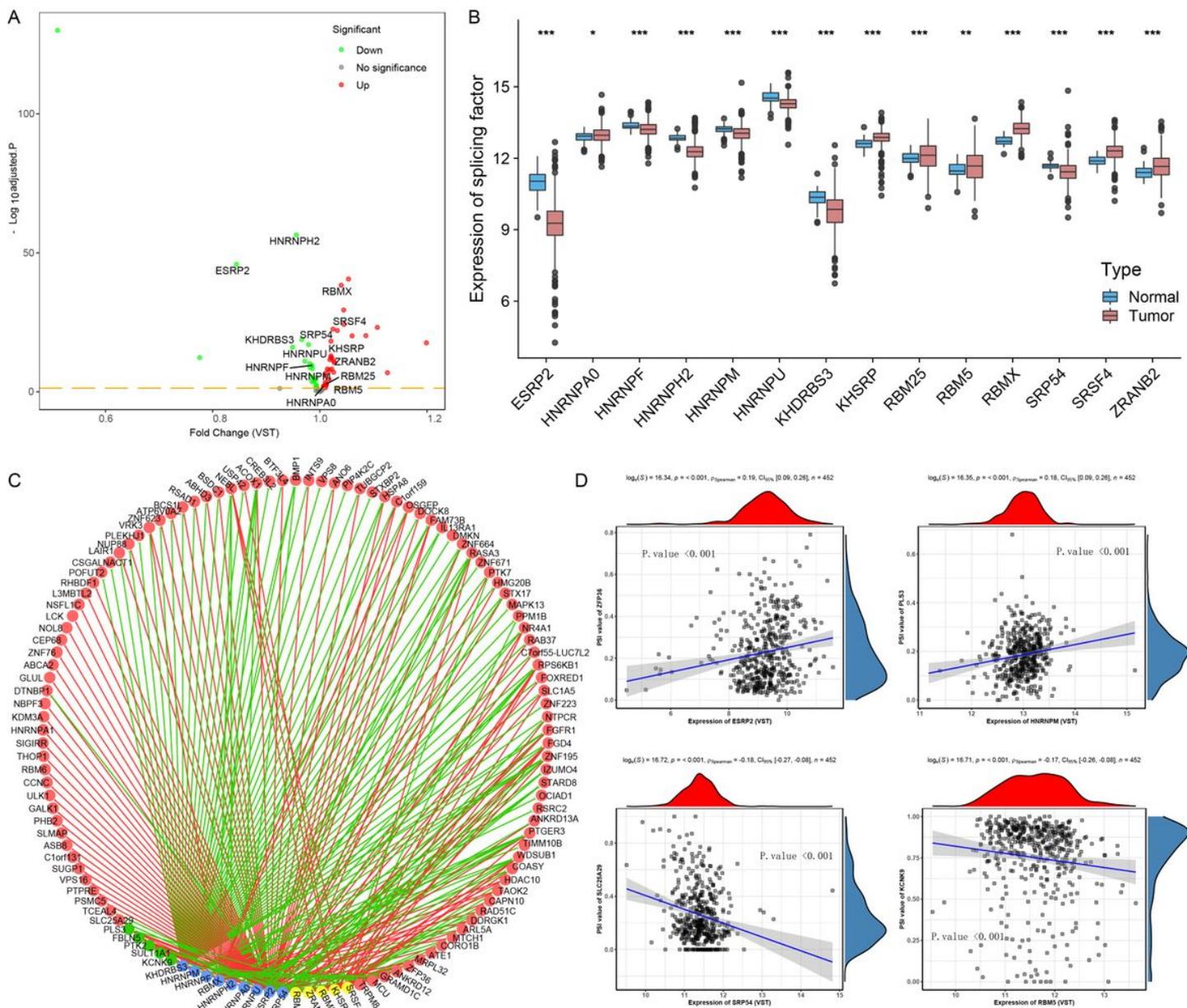


Figure 9

Aberrant prognostic splicing factors (SFs) and SF-AS regulatory network in ccRCC. (a) Volcano plot showing the differentially expressed SFs between tumor and paracancerous tissues. The red points indicate the upregulated SFs in tumor samples, whereas the green points represent the downregulated SFs in tumor tissues. The gene symbols of survival-associated ones are further labeled in the plot. (b) The difference of relative expression levels of aberrant prognostic SFs between ccRCC samples and normal samples. Boxes encompass 1st-3rd interquartile range of expression values, black lines in boxes represent the median values, whiskers indicate 1.5 times the interquartile range, and the black points indicate the outliers. (* p-value < 0.05, ** p-value < 0.01 and *** p-value < 0.001). (c) Splicing correlation network, showing the positive (red line) or negative (green line) relationship between expression level of

these SFs (yellow node indicates the risky one, while blue node indicates the protective one) and PSI value of most significant AS events (red node represents the adverse prognosis events, while green node represents favorable prognosis events). (d) Representative correlation pairs in SF-AS regulatory network.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement13.docx](#)
- [supplement14.docx](#)
- [supplement15.tif](#)
- [supplement16.tif](#)
- [supplement17.tif](#)
- [supplement18.tif](#)
- [supplement19.tif](#)
- [supplement20.tif](#)
- [supplement21.tif](#)