Alternative splicing perturbation landscape identifies RNA binding proteins as potential therapeutic targets in cancer

Junyi Li, Tao Pan, Liuxin Chen, Qi Wang, Zhenghong Chang, Weiwei Zhou, Xinhui Li, Gang Xu, Xia Li, Yongsheng Li, and Yunpeng Zhang

Alternative splicing (AS) plays an important role in gene regulation, and AS perturbations are frequently observed in cancer. RNA binding protein (RBP) is one of the molecular determinants of AS, and perturbations in RBP-gene network activity are causally associated with cancer development. Here, we performed a systematic analysis to characterize the perturbations in AS events across 18 cancer types. We showed that AS alterations were prevalent in cancer and involved in cancer-related pathways. Given that the extent of AS perturbation was associated with disease severity, we proposed a computational pipeline to identify RBP regulators. Pan-cancer analysis identified a number of conserved RBP regulators, which play important roles in regulating AS of genes involved in cancer hallmark pathways. Our application analysis revealed that the expression of 68 RBP regulators helped in cancer subtyping. Specifically, we identified four subtypes of kidney cancer with differences in cancer hallmark pathway activities and prognosis. Finally, we identified the small molecules that can potentially target the RBP genes and suggested potential candidates for cancer therapy. In summary, our comprehensive AS perturbation landscape analysis identified RBPs as potential therapeutic targets in cancer and provided novel insights into the regulatory functions of RBPs in cancer.

INTRODUCTION

Alternative splicing (AS) plays critical roles in post-transcriptional regulation of gene expression and contributes to protein complexity. Given the importance of AS events in gene regulation, perturbation of AS has been revealed in various types of cancer and it has been demonstrated that AS contributes to various cancer signaling pathways. Moreover, clinical relevance of AS events in cancer tissues has been demonstrated. In distinct cells or cancers, a precise splicing process controls cell decision and different cancer phenotypes. Although AS events in several candidate genes have been identified to contribute to cancer development, cancer-specific or recurrent splicing patterns on a transcriptome-wide scale have been less investigated.

Considerable efforts have been made to identify the perturbed AS events (PASEs) in cancer. However, the mechanisms by which molecular regulators perturb splicing remain unknown. Identification of these molecular regulators that perturb AS is critical for developing cancer biomarkers and therapeutic targets. With the development of computational and experimental technologies, both genetic and epigenetic determinants of AS events have been revealed in cancer. Deregulation of AS is potentially regulated by somatic mutations, antisense RNAs, RNA binding proteins (RBPs), transcriptional factors (TFs), and epigenetic factors. The PASEs can recapitulate the complex phenotype linked to mutations or expression dysregulation of RBPs. Our recent study also found that the somatic mutations potentially regulating AS events were enriched in RBPs. Widespread alterations of RBP genes have been found across cancer types. However, there is still a lack of computational or experimental methods to comprehensively identify the RBP regulators in cancer.

Multiple lines of evidence have demonstrated that RBPs play an indispensable role in cancer development and progression. Increasing numbers of RBPs are being identified as regulators of AS events in various types of cancer. For example, RBP RALY promotes the expression of alternatively spliced isoform v2 of arginine methyltransferase 1 and metastatic potential in breast cancer. RBP ZMAT3 is a regulator that inhibits splicing of CD44 in colorectal carcinoma. Thus, RBP-based cancer subtyping and targeting RBPs for cancer therapy have gained more attention. Our recent study also found that EIF2S2 knockdown combined with oxaliplatin treatment could be a potential combination therapy in cancer.
identification of potential drugs that target RBPs will provide novel insights for improving cancer therapy.

In this study, we systematically analyzed the genome-wide alternative splicing perturbations across cancer types. We found that there were prevalent changes in AS across cancers and PASEs were associated with disease severity. A computational pipeline was proposed to identify the RBP regulators of AS perturbation, and we found that these RBP regulators play important roles in cancer hallmark-related pathways. Moreover, we identified different subtypes based on the expression of RBP regulators, and patients in these subtypes showed distinct survival and pathway activities. Finally, we identified the drug-RBP correlations, which suggested new potential drugs for cancer therapy. Together, our integrative analysis of the AS perturbation landscape identified RBPs as potential therapeutic targets in cancer.

RESULTS

Landscape of perturbed AS events in cancer

To determine the global dysregulation of AS across cancer types, we systematically analyzed the mRNA splicing profiles in 18 human cancers from the The Cancer Genome Atlas (TCGA) cohort (Figure 1A; Table S1). There were in total 7,245 cancer samples and 735 normal samples. Throughout all cancer types, we found a substantial number of splicing events, and the proportions of splice types did not vary much between different cancer types (Figure 1B). In total, seven types of AS events were considered in our analysis, including exon skipping (ES), alternative donor site (AD), alternative acceptor site (AA), retained intron (RI), mutually exclusive exons (ME), alternative terminator (AT), and alternative promoter (AP). The most common splice event was ES, which accounted for 34.5%–41.4% of total events across cancers (Figure 1B: Table S2). In contrast, the least common AS event was ME, which accounted for only 0.3%–0.6% of total events across cancer types.
We next identified the PASEs in each cancer type. When directly contrasted to normal tissues in each cancer, we found that a high proportion of AS events was perturbed in cancer, ranging from 6% in esophageal carcinoma (ESCA) to 29% in kidney renal clear cell carcinoma (KIRC) (Table S3). The numbers of PASEs varied across cancer types (Figure 1C, top). We defined AS events with higher percent spliced-in (PSI) values as upregulated AS events in cancer, and those with lower PSI values were designated as downregulated AS events. We identified more upregulated AS events in KIRC and stomach adenocarcinoma (STAD; Figure 1C). There were more downregulated AS events in breast invasive carcinoma (BRCA), glioblastoma multiforme (GBM), and thyroid cancer (THCA; Figure 1C). Next, we explored the extent to which the different AS types were perturbed in cancer. Therefore, we calculated the odds ratio (OR) for each type of AS events. Although there was a higher number of ES events perturbed (Figure S1), the PASEs were likely to be enriched in AP and AT types (Figure 1C; Table S4, bottom). In addition, the upregulated PASEs in cancer were also enriched in RI splice type across cancers (Figure 1C; Table S4, bottom). This was consistent with the recent observation of RI as a hallmark of cancer stemness and aggressiveness.20 Taken together, these results suggest that there are prevalent AS alterations across cancer types.

PASEs imply tissue origin of cancers

It has been indicated that cancer types with similar tissue origins often share a number of molecular features, such as gene expression and DNA methylation.21,22 We next investigated whether cancer types with similar tissue origins exhibit similar PASEs. We first visualized the AS diversity across the cancer cohort with t-distributed stochastic neighbor embedding (t-SNE). This highlighted both the tissue-specific nature of AS events and cancer type-specific differences and commonalities (Figure 2A). We found that cancer types of similar origins and often described with similar molecular characteristics, such as kidney renal papillary cell carcinoma (KIRP), KIRC, and kidney chromophobe (KICH), liver hepatocellular carcinoma (LIHC) and cholangiocarcinoma (CHOL), or colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ), were commonly clustered closely together (Figure 2A). Moreover, we calculated the similarity between two cancers as the Jaccard index of the PASEs in two cancers (more details provided in Materials and methods). We observed the similar patterns based on a clustering of the similarity calculated as the proportion of shared PASEs (Figure 2B). These results are consistent with previous studies that focused on expression and methylation analyses23,24 and suggest that related AS mechanisms might operate in cancer types with similar tissue origin.

Next, we investigated the extent to which PASEs contribute to cancer specificity. We found that although a large number of PASEs showed cancer specificity, a number of PASEs were involved in numerous cancer types (Figure 2C). We defined these AS events as pan-cancer AS events. These pan-cancer differential AS events were involved in a number of well-known oncogenes (e.g., NFE2L2, RARA, and SGK1), tumor suppressor genes (e.g., CASP8, DNMT3A, FANCD2, and SMARCA4), and fusion genes (e.g., EIF4A2, SEPT6, and TCF12).25 Moreover, we found that the majority of these pan-PASEs were associated with patients’ survival and they exhibited similarly to various cancer types (Figure S2). We next performed functional enrichment analysis for these pan-cancer genes. We found that pan-cancer PASE-involved genes were enriched in a number of cancer-associated functional categories (Figure 2D; Table S5), such as translational initiation,26 protein binding, and cadherin binding involved in cell-cell adhesion.27 In addition, these genes exhibited high tissue-specific expression in testis, K562, and colorectal adenocarcinoma (Figure 2D). Taken together, these results imply that splicing abnormalities impact cancer biology by regulating cancer-related pathways.

Perturbed AS events are associated with disease severity

To determine the molecular players associated with the perturbed PASEs in cancer, we devised a versatile and comprehensive workflow to identify the RBP regulators (Figure 3A; see details in Materials and methods). Briefly, we first constructed the AS activity perturbation profiles for each patient by comparing the PSI to normal samples. Two activity scores (activated activity [SA] and repressed activity [SR]) were calculated by summarizing the perturbed level of each AS event in the same patient. Second, we identified the RBPs showing transcriptomic perturbation in cancer. RBP regulators were identified by correlating the expression of RBPs with the AS activity scores (Spearman correlation coefficient [SCC] p-adjusted < 0.1).

In total, we manually curated 2,038 RBPs and further classified them into nine clusters (Figure 3B). There were 139 ribosomal proteins, 636 mRNAs, and 89 noncoding RNAs (ncRNAs). However, 40.9% of RBPs had not been classified into specific clusters in the literature, suggesting that we still lack sufficient knowledge about their functions. We next explored the correlation between RBP expression and AS perturbation scores. We hypothesized that patients with higher AS perturbed scores (SA or SR) were likely to also show higher disease severity (Figure 3C). To test this hypothesis, we compared the survival rates among patients with different AS perturbed scores. We found that patients with higher SA or higher SR were likely to exhibit worse survival across cancer types, ~66.7% or 77.8% of cancer types with hazard ratio (HR) > 1 (Figure S3). In particular, patients with higher SA or SR in KIRC and LIHC had significantly worse survival (Figure 3D; log rank p values < 0.01). Moreover, we performed multivariate analysis by integrating age, sex, and stage. We found that the level of AS perturbation was an independent risk factor in KIRC and LIHC (Figures S4 and S5). Together, these results suggest that AS perturbations are associated with disease severity, where cancer patients with prevalent PASEs exhibit worse survival.

Identification of perturbed AS-related RBP regulators

Since we demonstrated that perturbed AS activity scores were associated with disease severity, we hypothesized that RBPs whose expression levels correlated with those scores were likely to regulate AS in cancer. First, we identified the differentially expressed RBPs in 18 cancer types. We found that more RBPs were upregulated in cancer (Figure 4A), which is consistent with recent studies.28,29 Next, we
calculated the number of cancer types in which these RBPs were differentially expressed. We found that more RBPs were only observed in one cancer type (Figure 4B), suggesting that different cancer types have distinct RBP dysregulation signatures. Moreover, there were some consistently dysregulated RBPs across the 18 cancer types analyzed. We identified 246 RBPs consistently up- or downregulated in >15 cancer types (Figure S6). Among these RBPs, we found that APOBEC3B, DCN, EZH2, and RAD51AP1 differentially expressed in all 18 cancer types (Table S6). APOBEC3B plays a key role in cancer and has been reported to be upregulated in many cancers.\(^{30,31}\) As a master regulator of chromatin regulation, EZH2 is overexpressed in human epithelial cancers.\(^{32}\) Moreover, we identified the differentially
Figure 3. AS perturbation correlated with disease severity
(A) The flowchart for identifying RBP regulators in cancer. (B) The proportion of curated RBP genes in different classes. (C) The cartoon shows the correlation between AS perturbed severity and survival rates. (D) Kaplan-Meier plots indicating survival of cancer patients with high and low AS activity scores. Top: SA and SR in KIRC. Bottom: SA and SR in LIHC.
expressed RBPs based on Wilcoxon rank-sum test. The majority of RBPs exhibited consistently differential expression (Table S7). All these RBPs provided a valuable resource for identifying the critical therapy targets in cancer.

We next identified the potential RBP regulators of AS events across cancer types based on the correlation between RBP expression and AS activity scores. In total, we identified 1,365 RBPs that were correlated with active scores (ActRBPs) and 1,045 RBPs correlated with repressive scores (ResRBPs) across 18 cancers (Figure S7). There were 211 ActRBPs and 91 ResRBPs observed in 15 or more cancer types (Figures 4C and 4D). The frequently dysregulated RBPs (APOBEC3B, DCN, EZH2, and RAD51AP1) were found in ActRBPs and identified as regulators in 18 cancer types (Table S8). Moreover, MEX3A, MRVI1, PUS7, FHL1, and SYNE1 were identified as the top ResRBPs, and they correlated with AS activity in 18 cancer types (Table S9). MEX3A is a stemness-related gene and plays an important role in cell proliferation, whereas FHL1 plays a tumor suppressor gene function in cancer and is repressed by epigenetic signals. Taken together, all these observations suggested that comprehensive analysis of the AS perturbation landscape can potentially capture the cancer-related RBPs in cancer.

**Function of pan-cancer RBP regulators**

We next calculated the number of cancer types for each RBP regulator. Many RBPs were only identified in one cancer type, although several RBPs were observed in >11 cancer types. Thus, we classified the RBP regulators into three types. The RBPs that occurred in only one cancer type were defined as cancer-specific RBPs, RBPs identified in 2–11 cancer types were defined as moderate RBPs, and those...
identified in >11 cancers were defined as pan-cancer RBPs (Figure 5A). In total, we identified 302 and 345 cancer-specific ActRBPs and ResRBPs, respectively. There were 71 and 22 pan-cancer ActRBPs and ResRBPs, respectively. These pan-cancer RBPs potentially regulated various types of AS events across cancer types (Figure 5B). In particular, ActRBP NME1 and ResRBP SYNE1 potentially regulate the highest number of AS events in cancer. NME1 is a metastasis suppressor that can potentially reduce metastasis without affecting primary tumor size; however, the precise mechanisms are unknown. Our results suggested that NME1 might regulate the AS of a number of cancer-related genes, such as CD44, CTNND1, and PTBP2. SYNE1 was found to have a higher mutation density. We found that this RBP can potentially regulate the AS events of numerous genes. These results suggest that identification of the potential AS targets provides insights into the function of RBP regulators.

To investigate the function of these pan-cancer RBPs, we next performed functional enrichment analysis based on their regulated AS event-involved genes. Here, we focused on 50 cancer hallmark-related pathways. At the false discovery rate (FDR) of <0.01, we identified 442 RBP-pathway associations among 93 RBPs and 16 pathways (Figure 5C). We found that all pan-cancer RBPs potentially regulated at least one cancer hallmark-related pathway. For these pathways, there were more RBPs involved in mitotic spindle, UV response, phosphatidylinositol 3-kinase (PI3K)-AKT-mammalian target of rapamycin (mTOR) signaling, and adipogenesis. Moreover, we found that RBPs played important roles in several metabolism-related pathways (Figure 5C). Indeed, increasing evidence has demonstrated their roles in metabolism. All these results suggest that pan-cancer RBPs play fundamental roles in cancer by regulating the AS events.

Patient stratification of kidney cancer based on RBPs

An increasing number of studies have demonstrated that AS events are able to discriminate between different subtypes of cancer and that they have prognostic impacts. We also found that the AS activity was associated with disease severity, particularly in kidney cancer. Thus, we next identified the RBP regulators and investigated
whether they could help in molecular subtyping of cancers. Here, we focused on two types of kidney cancer (KIRC and KIRP), as they exhibited similar results in our current analysis and previous pan-cancer results.23,46 We identified 68 common RBP regulators from these two cancer types (Figure 6A). Next, we grouped kidney cancer patients into different subtypes based on the expression of these RBPs. Consensus clustering was performed with k from 2 to 10.47 This analysis revealed that there were four subtypes in kidney cancer (Figure 6B). There were 298, 76, 426, and 26 patients in C1, C2, C3, and C4 subtypes, respectively (Table S10). We first investigated the source of these patients and found that the majority of patients for C1, C2, and C4 were KIRC. However, C3 subtype comprised both KIRC and KIRP patients (Figure S8).

Moreover, we explored the clinical information of patients from different subtypes. We found that the patients in C4 were diagnosed...
at a significantly younger age (Figure 6C; \( p < 0.001 \), Wilcoxon rank-sum test). The patients in C2 subtype had the highest age among all subtypes. We also calculated the proportion of patients in high stages (iii and iv) and low stages (i and ii). There were more high-stage patients in C2 and C4 than in C1 and C3 (Figure 6D; Fisher’s exact test \( p = 1.63E^{-3} \)). These results suggested that patients in C4 might be more severe, so we further investigated their survival rates. Indeed, we found that patients in C4 had the worst survival (Figure 6E; log rank test, \( p = 0.01 \)). The patients in C2 also had better survival than those in C4 (log rank test, \( p = 0.05 \)). These results suggest that the common RBP regulators are helpful in identifying the molecular subtypes of cancer.

**Distinct cancer pathways of kidney cancer subtypes**

We next investigated the extent to which pathway activities differ across different kidney cancer subtypes. First, we calculated the pathway activity for 50 cancer hallmark-related pathways in all patients based on gene set variation analysis (GSVA). Next, the differences of pathway activities were checked by analysis of variance (ANOVA). In total, we found that 80% (40/50) of these pathways showed different pathway activities across C1–C4 subtypes at the threshold FDR of <0.01 (Figure 7A). In particular, we found that patients in C2 and C4 had significantly higher complement pathway activities than those in C1 and C3 subtypes (Figure 7B; \( p \) values < 0.001, Wilcoxon rank-sum test). The complement system is a collection of serum proteins that are integral to inflammatory processes and innate immunity against infection.48 Several studies have demonstrated that some components generated by complement pathway activation could promote tumor cell growth and progression.49,50 These results are consistent with the worse survival of patients in C2 and C4 subtypes. Although complement pathway activities were similar between patients in C1 and C3, and between C2 and C4, there were significant survival differences among these pairs of groups (Figure 6E). We found that the bile acid metabolism pathways showed greater activities in C1 and C2 subtypes (Figure 7C; \( p \) values < 0.001, Wilcoxon rank-sum test), which might be a potential pathway indicator for distinguishing C1 from C3 and C2 from C4 subtypes. Bile acids play a critical role in the regulation of glucose, lipid, and energy metabolism.51

Moreover, we investigated the immune cell infiltration levels in patients. Those in C1 showed higher immune cell infiltration levels (Figure S9). Next, we explored the mutations in patients among different
subtypes. In total, five genes were identified with significant mutation frequency among subtypes (Fisher’s exact test p < 0.01; Figure S10). These genes have been found to play critical roles in cancer, such as VHL, HIF1A, and PBRM1. In addition, we investigated the extent to which these genes have been found to play critical roles in various types of cancer.56,57

With the development of high-throughput sequencing, a number of RBPs can affect many characteristics of cancer, which suggests that RBPs might be a good therapeutic target for cancer treatment.58 Next, to identify the potential RBP candidates for cancer therapy, we evaluated the correlation between the expression levels of RBPs and drug activities across cancer types. We found that ~1,088 RBPs correlated with the drug half-maximal inhibitory concentration (IC50) values across 10 cancers (Figure 8A; absolute values of correlation coefficient |RI| > 0.5 and p values < 0.01). We next focused on the drug-RBP correlations in KIRC, and these correlated RBPs and drugs formed a dense network (Figure 8B). Specifically, we found that several drugs, such as pemetrexed, masitinib, and omipalisib, correlated with more RBPs. Pelitinib is a potent irreversible epidermal growth factor receptor-tyrosine kinase inhibitor currently in clinical trials for the treatment of cancer.55 Moreover, several RBPs, such as ATP5F1, BCAS2, and RRP12, also correlated with more drugs. These RBPs have been demonstrated to play key roles in various types of cancer.56,57

In addition, we extracted the sub-network for 68 RBPs that were used for kidney cancer subtyping. Ten drugs potentially correlated with the expression of five RBPs (Figure 8C), including SNRPD1, RNPS1, PP1A, MRPL14, and PP1H. The expression of PP1H correlated with masitinib and QL-XII-47. Next, we computationally modeled the altered expression of RBPs by dividing cell lines into two groups. Then, we compared the IC50 values between two groups with different RB expression and found significant inter-group differences (Figure S13). These results suggest that alteration of RBP expression might affect the IC50 of cell lines. Moreover, high expression of PP1H was significantly associated with worse overall survival and disease-free survival in kidney cancer (log rank test, p value = 0.0066 and p value = 0.013, respectively; Figure 8D). All these results suggest that these two drugs might be potentially used for cancer therapy.

DISCUSSION

There is accumulating evidence that RBPs regulate the expression and AS of thousands of transcripts, and more and more RBPs have been reported to be involved in various types of cancer. In the present study, we integrated the AS perturbation and genome-wide expression of RBPs and proposed a computational method to identify the RBP regulators in cancer. Our hypothesis was that RBPs can regulate AS in cancer. These RBPs also exhibited dysregulated expression in cancer and correlated with cancer-related pathway activities. Our analyses demonstrated that these RBP regulators play important roles in cancer hallmark-related pathways and may be helpful in cancer subtyping and identification of potential drug targets. All these results provided new insights into the RBP functions in cancer as well as cancer therapy. Moreover, we computationally analyzed the perturbation of AS across cancer types. We identified the important RBPs correlated with cancer pathway activities. However, these prioritized ASs and RBPs need to be further functionally validated in cell lines or animal models.

Therapeutic targets of RBPs in cancer

As shown in previous studies and our current observations, deregulated RBPs can affect many characteristics of cancer, which suggests that RBPs might be a good therapeutic target for cancer treatment.59 Next, to identify the potential RBP candidates for cancer therapy, we evaluated the correlation between the expression levels of RBPs and drug activities across cancer types. We found that ~1,088 RBPs correlated with the drug half-maximal inhibitory concentration (IC50) values across 10 cancers (Figure 8A; absolute values of correlation coefficient |RI| > 0.5 and p values < 0.01). We next focused on the drug-RBP correlations in KIRC, and these correlated RBPs and drugs formed a dense network (Figure 8B). Specifically, we found that several drugs, such as pemetrexed, masitinib, and omipalisib, correlated with more RBPs. Pelitinib is a potent irreversible epidermal growth factor receptor-tyrosine kinase inhibitor currently in clinical trials for the treatment of cancer.55 Moreover, several RBPs, such as ATP5F1, BCAS2, and RRP12, also correlated with more drugs. These RBPs have been demonstrated to play key roles in various types of cancer.56,57

In addition, we extracted the sub-network for 68 RBPs that were used for kidney cancer subtyping. Ten drugs potentially correlated with the expression of five RBPs (Figure 8C), including SNRPD1, RNPS1, PP1A, MRPL14, and PP1H. The expression of PP1H correlated with masitinib and QL-XII-47. Next, we computationally modeled the altered expression of RBPs by dividing cell lines into two groups. Then, we compared the IC50 values between two groups with different RB expression and found significant inter-group differences (Figure S13). These results suggest that alteration of RBP expression might affect the IC50 of cell lines. Moreover, high expression of PP1H was significantly associated with worse overall survival and disease-free survival in kidney cancer (log rank test, p value = 0.0066 and p value = 0.013, respectively; Figure 8D). All these results suggest that these two drugs might be potentially used for cancer therapy.

With the development of high-throughput sequencing, a number of RBP-gene regulations were identified. However, only 40 RNA binding domains (RBDs) of RBPs have been identified to date, although their number is constantly increasing.58 The presence of multiple RBDs in a RBP enables a more specific and higher-affinity interaction of the RBP with its targets. Thus, we need to analyze the regulation among RBPs and genes in a network-based view. Recently developed cross-linking and immunoprecipitation (CLIP)-based methods have been employed with demonstrable success in the transcriptome-wide identification of binding sites for RBPs.59 However, additional work is still needed to establish protocols that would allow obtaining quantitative data on RBP-RNA interactions. Identification of the RBP targets will help in understanding their roles in cancer.

Moreover, immunotherapy is now becoming a frontline treatment for many tumor types. A few notable examples of AS events have clearly been demonstrated to regulate T cell responses to antigens.60 Additionally, a growing number of RBPs were found to be involved in cancer immunology. RBP PTB1P1 was found to be important for the progression of B cells through late S phase of the cell cycle and for affinity maturation.61 AT-rich interactive domain 5A (ARID5A/Arid5a) has recently been recognized to be associated with inflammatory autoimmune diseases.62 Distinct isoforms of RBPs have been found to play specific antiviral and immune resolution functions.63 Thus, systematical understanding of the consequences of AS and RBP regulation in the adaptive immune system and cancer immunology will give us a new view of functions of AS in cancer.

In summary, our results provided a detailed landscape for AS perturbation across cancer types and identified the potential RBP regulators. Functional analyses and drug screening of RBPs may be helpful to pave the way for the development of cancer therapies.

MATERIALS AND METHODS

Genome-wide AS profiling across human cancers

To systematically investigate the AS landscape across cancer, we first retrieved the genome-wide AS data in 10,699 samples representing 33 cancer types.64 The AS profiles of patients were downloaded from the TCGA SpliceSeq database. The transcript splicing patterns and splicing events were identified by the SpliceSeq pipeline.65 The
splicing events were quantified by the PSI value, which was defined as the number of reads that a transcript element was presented divided by the total number of reads covering the AS event. In total, seven types of AS events were considered, including ES, AD, AA, RI, ME, AT, and AP.

Identification of perturbed AS events in cancer

We identified the PASEs in 18 cancer types with more than five normal samples. Wilcoxon rank-sum test was used to identify the differential AS events in each cancer type. The p values were corrected by Benjamini-Hochberg (BH) method. The AS events with >1.2-fold changes and adjusted p values < 0.05 were identified as perturbed.

In addition, we calculated the OR for each type of AS event to explore which AS types were more likely to be perturbed in cancer. OR is a statistic that is similar to risk ratio (RR). The OR is the ratio of the odds of an AS event occurring in differential group to the odds of it occurring in total AS events. An OR >1 indicates that the AS event...
is more likely to occur in the differential group. The OR for each AS type \( m \) in cancer \( n \) was calculated by

\[
OR_{mn} = \frac{D_{mn}/A_{mn}}{D_n/A_n}, \quad m = 1, 2, \ldots, 7; \quad n = 1, 2, \ldots, 18,
\]

where \( D_{mn} \) is the number of perturbed AS type \( m \) in cancer \( n \), \( A_{mn} \) is the total number of AS type \( m \) in cancer \( n \), \( D_n \) is the total number of PASEs in cancer \( n \), and \( A_n \) is the total number of AS events observed in cancer \( n \). Moreover, we calculated the cancer similarity based on the shared PASEs, similar to our previous method.\(^8\)

The t-distributed stochastic neighbor embedding of cancer patients

We first used the “var” function in Python to identify the AS events that exhibited high variation of PSI values across tumors. All the AS events exhibiting difference in at least one cancer type were considered in our analysis. The 100 top-ranked AS events were used, and all tumor samples were modeled by a two-dimensional point.

Manual collection and classification of human RBPs

We manually curated the recently published literature to collect comprehensive RBPs in human. First, we obtained three human RBP gene lists from the literature, including 1,350,\(^6\) 799,\(^7\) and 1,072\(^7\) genes. Next, 1,148 RBPs in mouse were retrieved from the Liao et al. study.\(^8\) We used BioMart in Ensembl to map these RBPs to human genes. Next, 1,550 genes annotated as “binding” and “splicing” in Gene Ontology were collected. In total, 2,038 RBP genes were obtained for analysis. These RBPs were classified into nine sub-classes based on literature annotation, including ribosomal-binding protein, mRNA-binding protein, pre-rRNA-binding protein, tRNA-binding protein, small nuclear RNA (snRNA)-binding protein, small nucleolar RNA ( snoRNA)-binding protein, ncRNA-binding protein, diverse-binding protein, and unknown.

Transcriptomic dysregulation analysis of RBPs in cancer

Genome-wide expressions of RBPs in 33 cancer types were obtained from the TCGA project. We first filtered the genes with expression level 0 in >30% of samples. Next, a t test was used to investigate the difference in expression of RBPs in cancer versus normal tissue. Here, 18 cancer types with more than five normal samples were considered. p values were adjusted by BH method, and RBPs with >1.2-fold changes and p-adjusted < 0.05 were identified as dysregulated RBPs in cancer. Moreover, Wilcoxon rank-sum test was used to identify dysregulated RBPs. In addition, another two independent kidney cancer transcriptomes were downloaded from Gene Expression Omnibus (GEO) under the accession numbers GSE29609 and GSE2748. These two datasets were combined, and bath effects were corrected by ComBat.

Identification of RBPs correlating with perturbed AS

To identify the RBPs that may potentially regulate the AS events in cancer, we proposed a three-step method that integrated RBP expression and AS profiling. First, we constructed the perturbed AS profiling for each patient in each cancer. For the specific AS event \( i \), the perturbed level in patient \( j \) was calculated as follows:

\[
\Delta PSI_{ij} = PSI_{ij} - \frac{\sum_{k=1}^{N} PSI_{kj}}{N}
\]

where \( N \) is the number of normal samples in specific cancer. Next, we calculated two activity scores (active and repressive scores, SA and SR) for all AS events across cancer patients. For each patient \( j \), these two scores were calculated as follows:

\[
\begin{align*}
SA(j) &= PSI(\text{active})_j = \sum_{a=1}^{p} \Delta PSI_{aj}; \quad \text{if } \Delta PSI_{aj} > 0 \\
SR(j) &= PSI(\text{repressive})_j = \sum_{b=1}^{q} \Delta PSI_{bj}; \quad \text{if } \Delta PSI_{bj} < 0
\end{align*}
\]

where \( p \) and \( q \) are the number of AS events with \( \Delta PSI \) greater or less than 0, respectively.

Second, we identified RBPs with transcriptomic perturbations in cancer by t test. We hypothesized that the expression levels of RBP regulators would be correlated with the activity scores. Thus, we calculated the SCC between the RBP expression level and active and repressive scores across cancer patients. p values of SCC < 0.1 were selected for further analysis. In addition, the labels of patients were perturbed, and we recalculated the SCC. This procedure was repeated 1,000 times for each RBP. The p value for random test was calculated as the number of SCCs in random conditions that were greater than real conditions. p values were adjusted by BH method, and we identified the RBPs with p-adjusted < 0.05. RBPs that correlated with active/repressive scores were defined as active/repressive RBP regulators, respectively.

Functional analysis of pan-cancer RBP regulators

To identify the function of RBPs, we performed functional enrichment analysis based on the AS event-involved genes. In our analysis, 50 cancer hallmark-related pathways were used.\(^9\) The hypergeometric test was used for this process, and the p values were calculated as follows:

\[
p(\text{RBP}_{i} \in \text{Path}_j) = 1 - F(N_j - 1\mid N, N_i, N_j) = 1 - \sum_{t=0}^{N_i-1} \binom{N_j}{t} \binom{N - N_j}{N_i - t}
\]

where \( N \) is the total number of human coding genes, \( N_j \) represents the AS genes regulated by RBP \( i \), \( N_i \) is the number of genes in pathway \( j \), and \( N_{ij} \) are the overlapping genes between RBP-regulated genes and those in pathway. All p values were subjected to FDR correction. All RBP-pathway pairs with FDR < 0.05 were identified and visualized by Cytoscape 3.7.1.\(^9\)
Identification of subtypes of cancer
The common RBP regulators that potentially regulated AS events in KIRC and KIRP were identified. Consensus ward linkage hierarchical clustering of the samples and RBPs identified the subtypes with the stability of the clustering increasing from $k = 2$ to $k = 10$. For each iteration process, we selected 80% of the kidney cancer samples, which was repeated 100 times. This process was performed by using the R package “ConsensusClusterPlus.”  

Pathway activities of cancer patients
To calculate the pathway activities for each cancer patient, we used GSVA and 50 cancer hallmark-related pathways from MSigDB. The pathway differences among patients in different subtypes were investigated by ANOVA. The p values were adjusted by BH method, and pathways with FDR $<0.01$ were identified.

Immune cell infiltration levels in cancer patients
We downloaded the immune cell infiltration levels from TIMER2.0 (http://timer.cistrome.org/). We used “kruskal.test” to evaluate the differences among different subtypes. The immune cell infiltration levels with p-adjusted $<0.01$ were shown by heatmap.

Identification of drug-related RBPs
We first downloaded the IC50 values for a number of drugs across cell lines from Genomics of Drug Sensitivity in Cancer (GDSC). These cell lines were classified into different cancer types. For each cancer type, we calculated the correlation coefficient between the expression of RBPs and IC50 values of a drug across cell lines. The drug-RBP pairs with $|R| > 0.5$ and p value $< 0.01$ were identified. The drug-RBP pairs were visualized by Cytoscape 3.7.1.

Survival analysis
The clinical information for cancers was obtained from the TCGA project via the TCGAbiolinks R package. The function survdiff in the survival R package was used for exploring the difference in survival time among different groups of patients.

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2021.04.005.

ACKNOWLEDGMENTS
This work was supported by Hainan Provincial Natural Science Foundation of China (grant no. 820MS053); the National Key R&D Program of China (grant no. 2018YFC2000100); the National Natural Science Foundation of China (grant nos. 61873075, 32060152, 32070673, 31970646, and 31871338); Natural Science Foundation for Distinguished Young Scholars of Heilongjiang Province (JQ2019C004); and Heilongjiang Touyan Innovation Team Program and University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (UNPYSCT-2016189 and UNPYSCT-2017059).

AUTHOR CONTRIBUTIONS
Y.L., Xia Li, and Y.Z. designed the research studies. J.L., I.C., and T.P. acquired the data and conducted the data analysis. Q.W., Z.C., W.Z., Xinhui Li, and G.X. designed the figures with input from Y.L., Y.L., Xia Li, and Y.Z. wrote the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS
The authors declare no competing interests.

REFERENCES
1. Blencowe, B.J. (2017). The Relationship between Alternative Splicing and Proteomic Complexity. Trends Biochem. Sci. 42, 407–408.
2. Koch, L. (2017). Alternative splicing: A thermometer controlling gene expression. Nat. Rev. Genet. 18, 515.
3. Kahles, A., Lehmann, K.V., Toussaint, N.C., Hüser, M., Stark, S.G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., and Rätsch, G. (2018). Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell 34, 211–224.e6.
4. Climente-González, H., Porta-Pardo, E., Godzik, A., and Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. Cell Rep. 20, 2215–2226.
5. Li, Y., McGurl, D.J., Xu, J., Mills, G.B., Sahni, N., and Yi, S. (2018). Gene Regulatory Network Perturbation by Genetic and Epigenetic Variation. Trends Biochem. Sci. 43, 576–592.
6. Zhang, Y., Yan, L., Zeng, J., Zhou, H., Liu, H., Yu, G., Yao, W., Chen, K., Ye, Z., and Xu, H. (2019). Pan-cancer analysis of clinical relevance of alternative splicing events in 31 human cancers. Oncogene 38, 6678–6695.
7. Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. 40, 1413–1415.
8. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476.
9. Jung, H., Lee, D., Lee, J., Park, D., Kim, Y.J., Park, W.Y., Hong, D., Park, P.J., and Lee, E. (2015). Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nat. Genet. 47, 1242–1248.
10. Li, Y., Sahni, N., Pancra, R., McGurl, D.J., Xu, J., Hua, X., Coulombe-Huntington, L., Ryan, M., Tychon, B., Sudhakar, D., et al. (2017). Revealing the Determinants of Widespread Alternative Splicing Perturbation in Cancer. Cell Rep. 21, 798–812.
11. Morrissy, A.S., Griffith, M., and Marra, M.A. (2011). Extensive relationship between antisense transcription and alternative splicing in the human genome. Genome Res. 21, 1203–1212.
12. Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P.M., Green, M.R., Riva, S., and Biamonti, G. (2005). Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. Mol. Cell 20, 881–890.
13. Han, H., Braunschweig, U., Gonatopoulos-Pournatzis, T., Weatheritt, R.J., Hirsch, C.L., Ha, K.C.H., Radovani, E., Nabeel-Shah, S., Sterne-Weiler, T., Wang, J., et al. (2017). Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. Mol. Cell 65, 539–553.e7.
14. Bechara, E.G., Sebestyén, E., Bernardos, I., Eyras, E., and Valcárcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. Mol. Cell 52, 720–733.
15. Sebestyén, E., Singh, B., Mitána, B., Pagés, A., Matteo, F., Pujana, M.A., Valcárcel, J., and Eyras, E. (2016). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Genome Res. 26, 732–744.
16. Bondy-Chorney, E., Baldwin, R.M., Didillon, A., Chabot, B., Jasmin, B.J., and Côté, J. (2017). RNA binding protein RALY promotes Protein Arginine Methyltransferase 1 alternatively spliced isoform v2 relative expression and metastatic potential in breast cancer cells. Int. J. Biochem. Cell Biol. 91 (Pt B), 124–135.
56. Brüggemann, M., Gromes, A.,Poss, M., Schmidt, D., Klümper, N., Tolkach, Y., Dietrich, D.,Kristiansen, G., Müller, S.C., and Ellinger, J. (2017). Systematic Analysis of the Expression of the Mitochondrial ATP Synthase (Complex V) Subunits in Clear Cell Renal Cell Carcinoma. Transl. Oncol. 10, 661–668.

57. Zhou, K.R., Liu, S., Sun, W.J., Zheng, L.L., Zhou, H., Yang, J.H., and Qu, L.H. (2017). ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. Nucleic Acids Res. 45 (D1), D43–D50.

58. Kishore, S., Luber, S., and Zavolan, M. (2010). Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. Brief. Funct. Genomics 9, 391–404.

59. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129–141.

60. Martinez, N.M., and Lynch, K.W. (2013). Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. Immunol. Rev. 253, 216–236.

61. Monzón-Casanova, E., Screen, M., Díaz-Muñoz, M.D., Coulson, R.M.R., Bell, S.E., Lamers, G., Solimena, M., Smith, C.W.J., and Turner, M. (2018). The RNA-binding protein PTBP1 is necessary for B cell selection in germinal centers. Nat. Immunol. 19, 267–278.

62. Nyati, K.K., Zaman, M.M., Sharma, P., and Kishimoto, T. (2020). Arid5a, an RNA-Binding Protein in Immune Regulation: RNA Stability, Inflammation, and Autoimmunity. Trends Immunol. 41, 255–268.

63. Schwerk, J., Soweg, F.W., Ryan, A.P., Thomas, K.R., Hatfield, L.D., Ozarkar, S., Forero, A., Kell, A.M., Roby, J.A., So, L., et al. (2019). RNA-binding protein isoforms ZAP-S and ZAP-I have distinct antiviral and immune resolution functions. Nat. Immunol. 20, 1610–1620.

64. Ryan, M., Wong, W.C., Brown, R., Akhani, R., Su, X., Broom, B., Melott, J., and Weinstein, J. (2016). TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. Nucleic Acids Res. 44 (D1), D1018–D1022.

65. Ryan, M.C., Cleland, J., Kim, R., Wong, W.C., and Weinstein, J.N. (2012). SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. Bioinformatics 28, 2385–2387.

66. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Series B 57, 289–300.

67. Sundararaman, B., Zhan, L., Blue, S.M., Stanton, R., Elkins, K., Olson, S., Wei, X., Van Nostrand, E.L., Pratt, G.A., Huelga, S.C., et al. (2016). Resources for the Comprehensive Discovery of Functional RNA Elements. Mol. Cell 61, 903–913.

68. Liao, Y., Castello, A., Fischer, B., Leicht, S., Föréhr, S., Frese, C.K., Ragan, C., Kurscheid, S., Pagler, E., Yang, H., et al. (2016). The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. Cell Rep. 15, 1456–1469.

69. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.

70. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14, 7.

71. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics 27, 1739–1740.

72. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beaz, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 41, D955–D961.

73. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Gardolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Biocductor package for integrative analysis of TCGA data. Nucleic Acids Res. 44, e71.