



# Screening and Identification of Survival-Associated Splicing Factors in Lung Squamous Cell Carcinoma

Min Chen<sup>1†</sup>, Rui Zhu<sup>1†</sup>, Fangzhou Zhang<sup>2,3\*</sup> and Liucun Zhu<sup>1\*</sup>

<sup>1</sup>School of Life Sciences, Shanghai University, Shanghai, China, <sup>2</sup>School of Materials Science and Engineering, Institute of Materials, Shanghai University, Shanghai, China, <sup>3</sup>Shaoxing Institute of Technology, Shanghai University, Shanghai, China

## OPEN ACCESS

### Edited by:

Quan Zou,  
University of Electronic Science and  
Technology of China, China

### Reviewed by:

Yingying Dong,  
Soochow University, China  
Xuebing Li,  
Tianjin Medical University General  
Hospital, China

### \*Correspondence:

Fangzhou Zhang  
zhangfzh@shu.edu.cn  
Liucun Zhu  
zhuliucun@shu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 28 October 2021

Accepted: 27 December 2021

Published: 20 January 2022

### Citation:

Chen M, Zhu R, Zhang F and Zhu L  
(2022) Screening and Identification of  
Survival-Associated Splicing Factors in  
Lung Squamous Cell Carcinoma.  
Front. Genet. 12:803606.  
doi: 10.3389/fgene.2021.803606

Lung squamous cell carcinoma (LUSC) is a disease with high morbidity and mortality. Many studies have shown that aberrant alternative splicing (AS) can lead to tumorigenesis, and splicing factors (SFs) serve as an important function during AS. In this research, we propose an analysis method based on synergy to screen key factors that regulate the initiation and progression of LUSC. We first screened alternative splicing events (ASEs) associated with survival in LUSC patients by bivariate Cox regression analysis. Then an association network consisting of OS-ASEs, SFs, and their targeting relationship was constructed to identify key SFs. Finally, 10 key SFs were selected in terms of degree centrality. The validation on TCGA and cross-platform GEO datasets showed that some SFs were significantly differentially expressed in cancer and paracancer tissues, and some of them were associated with prognosis, indicating that our method is valid and accurate. It is expected that our method would be applied to a wide range of research fields and provide new insights in the future.

**Keywords:** lung squamous cell carcinoma, alternative splicing, splicing factor, bivariate cox regression, bipartite graph

## INTRODUCTION

Lung cancer is one of the most common malignant tumors, and about 85% of cases are non-small cell lung cancer (NSCLC) (Wang et al., 2019). According to pathological classification, NSCLC can be divided into lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD) (Cheng et al., 2019). Compared with LUAD, patients with LUSC have a poorer treatment outcome and prognosis (Li et al., 2018). In recent years, targeted therapies for specific genes have greatly improved the living conditions of patients with advanced LUAD. However, LUSC patients respond poorly to targeted therapies due to the lack of driver mutations, and the specific molecular mechanisms of LUSC pathogenesis and progression have not been systematically assessed. As a result, further exploration of the molecular mechanisms underlying the development of LUSC is essential for the development of more effective therapeutic regimens.

Alternative splicing (AS) is an important post-transcriptional regulatory mechanism. A single gene can generate more than one mRNA transcript through AS, and each mRNA transcript encodes a protein with a different structure and function (Baralle and Giudice, 2017). More than 95% of human genes experience AS under normal physiological conditions. On the one hand, the AS process regulates the tissue-specific and stage-specific expressions of specific genes during human development (Xu et al., 2002; Pan et al., 2008) and is essential for normal biological processes, such as hematopoiesis (Wong et al., 2018), brain development (Matsuda et al., 2019), and muscle

function (Nakka et al., 2018). On the other hand, abnormal AS triggers a series of tumor-related processes, including cell proliferation (Xie et al., 2019), apoptosis (Tyson-Capper and Gautrey, 2018), epithelial-mesenchymal transition (EMT) (Pradella et al., 2017), and tumor invasion and metastasis (Chen et al., 2017; Wang et al., 2017) in response to hypoxia (Han et al., 2017), thereby promoting malignant cell transformation and providing a survival advantage (Climente-González et al., 2017; Moncada et al., 2020). The AS process is regulated by splicing factors (SFs), and abnormal expression of SFs is the main contributor to overall changes in alternative splicing events (ASEs) in malignancies (David and Manley, 2010; Dvinge et al., 2016; Su et al., 2018). Therefore, exploring abnormal ASEs and SFs in malignant tumors may provide new insights into the mechanisms of tumorigenesis and progression.

Recent studies have paid more attention to assessing the clinical significance of ASEs and SFs in cancers and their potential pathogenic pathways and regulatory networks. The abnormal ASEs and SFs, which make network dysregulated, have been shown to modulate malignant transformation of cells and epithelial-mesenchymal transition (Sveen et al., 2016). Several excellent studies have also discussed the role of SFs in DNA damage (Shkreta and Chabot, 2015) or in carcinogenesis and anticancer therapies (Miura et al., 2012; Shkreta et al., 2013). However, SFs have the potential to become molecular markers and therapeutic targets for malignancies (Anczukow and Krainer, 2016; Yuan et al., 2017; Park et al., 2019). Although there is an increasing systematic analysis of AS signatures and the effect of SFs in colorectal cancer, glioblastoma, breast cancer, and ovarian cancer (Dorman et al., 2014; Suo et al., 2015; Zong et al., 2018), the analytical methods for identifying tumor-associated SFs remain deficient. Only univariate difference and survival analysis were performed in these studies (Zhu et al., 2018; Hu et al., 2019; Zhao et al., 2020). However, biological processes are complex and are mostly regulated by multiple factors rather than a single factor. It is indicated that, as a whole, some factors would have a high correlation with the tumor process, but this would show a low correlation when they are separated. Hence, we propose an analysis method based on synergy to screen key factors that regulate the initiation and progression of LUSC. We first screened the ASEs associated with overall survival (OS-ASEs) from combinations consisting of two ASEs using bivariate Cox regression and AUROC. Then an association network consisting of OS-ASEs, SFs, and their targeting relationship was constructed to identify key SFs. This method can screen a relatively complete set of OS-ASEs to a certain extent, thereby improving the completeness for subsequent screening of key SFs and providing new ideas for LUSC mechanism research.

## MATERIALS AND METHODS

### Data Collection and Preprocessing

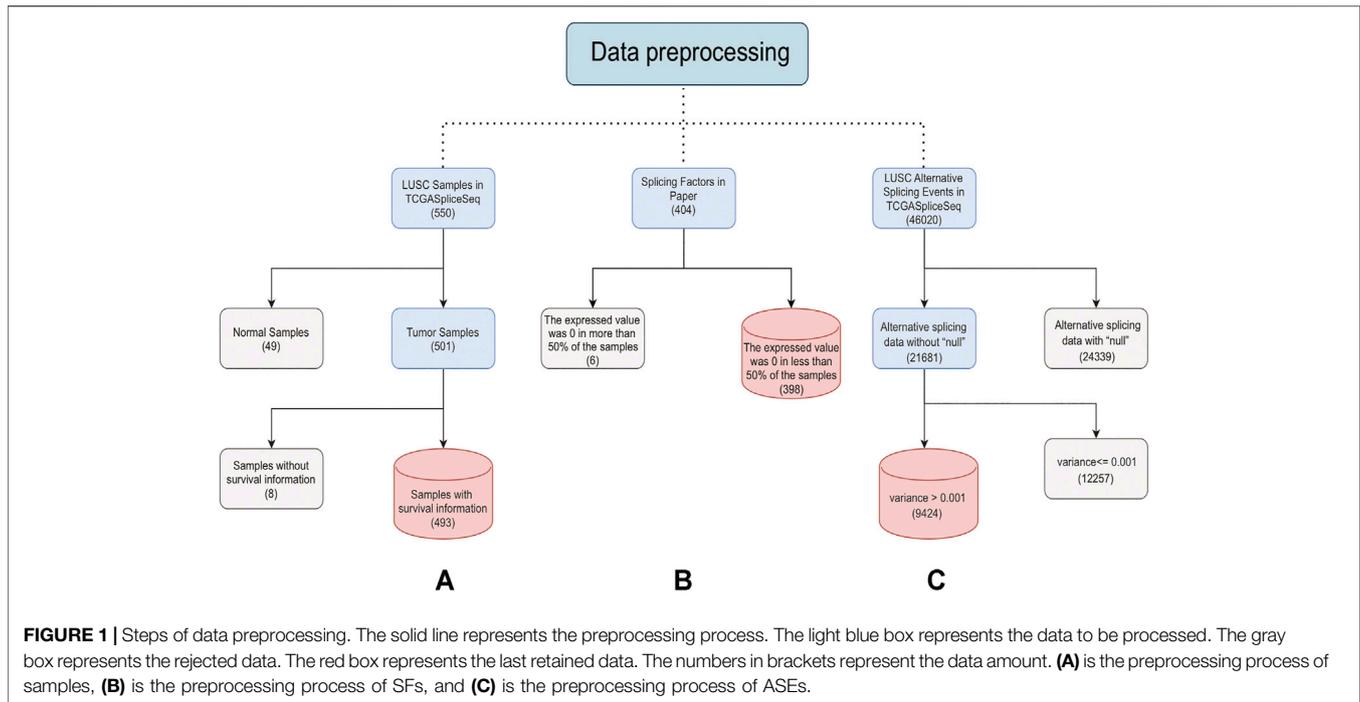
Clinical information and expression levels of LUSC patients (generated by RNA-seq) were collected from The Cancer Genome Atlas (TCGA) database. Additionally, ASEs data were

retrieved from the TCGASpliceSeq database (Ryan et al., 2016). In TCGASpliceSeq, the Percent Spliced In (PSI) values are computed for each possible splice event in each gene. PSI is the ratio of reads indicating the presence of a transcript element versus the total reads covering the event. The cross-platform validation set, including GSE157010, GSE3268, and GSE6044 (**Supplementary Table S1**), was downloaded from the NCBI-GEO database (Barrett et al., 2013). SFs are protein factors involved in the splicing process of pre-RNA. A total of 404 SFs were collected in this study (Wu et al., 2020), as shown in **Supplementary Table S2**.

The TCGA database included 550 LUSC samples, 501 of which were tumor samples. After removing 8 samples with no clinical information, 493 tumor samples were retained for subsequent analysis (**Supplementary Table S3**). The TCGASpliceSeq database contained a total of 46,020 ASEs for LUSC, of which 9424 ASEs were retained for subsequent analysis by removing ASE containing “null” and then excluding ASEs with variances less than 0.001 in all samples (**Supplementary Source Code S1**) (**Supplementary Source Code S2**). The distinguishable visualization UpSet plot, generated by *UpSetR* (version 1.4.0) (Wang et al., 2021), was used to quantitatively analyze the intersections among the seven types of ASEs in LUSC. The expressions of 404 SFs were extracted after being normalized by log<sub>2</sub>(FPKM+1) (Bullard et al., 2010). SFs with expression values of 0 in half of the samples were excluded, and 398 SFs were finally retained for subsequent analysis (**Figure 1**). The GSE157010 dataset constitutes 235 LUSC tumor samples, each containing clinical information. The GSE3268 dataset represents 5 tumor samples from LUSC patients and paired normal samples. The GSE6044 dataset includes 5 normal samples and 15 tumor samples. Ten of these 15 tumor patients have not received platinum-based therapy, and the other five have. Probe IDs for each GEO dataset were converted to Ensembl ID. When multiple probes correspond to an Ensembl ID, only the probe with the highest mean is retained. The batch correction was performed to eliminate the batch effect of three datasets using *normalizeBetweenArrays* function of *limma* (version 3.46.0).

### Methods for Screening Alternative Splicing Events Associated With Overall Survival

In order to investigate the prognostic value of ASEs in LUSC patients, all bivariate ASEs combinations were first constructed. Then Cox proportional risk hypothesis tests and bivariate Cox proportional risk regressions were performed using the survival package in R (Bradburn et al., 2003). The significance of the independent variables in the regressions was tested using likelihood ratio tests (Hazra and Gogtay, 2017). Additionally, the area under the receiver operating characteristic curve (AUROC) was used to show the sensitivity and specificity of the bivariate combination model in predicting OS (Linden, 2006). Values greater than 0.8 were considered excellent combinations. The two indicators mentioned above, the *p*-value of the likelihood ratio test and the AUROC, were used to screen OS-ASEs.



## Methods for Association Network Construction and Analysis

Spearman correlation analysis was performed to explore the correlation between the PSI values of OS-ASEs and the expression levels of SF genes (Bishara and Hittner, 2012). The correlation networks visualization was visualized by *Eveen* (Chen T. et al., 2021) with SFs and the OS-ASEs as vertices and the Spearman significant correlation between them as edges. It is assumed that the value of a vertex in a network depends first on its position in the network. More central vertex indicates a greater impact on the structure and function of the network (Kitsak et al., 2010). The importance of a vertex in the network is usually expressed by degree centrality, which is the number of connected edges of the vertex in the network (Freeman, 1978).

## Validation Methods for Alternative Splicing Events and Splicing Factors Functions

In order to identify potential mechanisms of OS-ASEs in LUSC, the survival-related genes were analyzed by Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, which were both done by *DAVID* (Huang et al., 2009). The results of KEGG analysis were presented by bubble plots generated by *ggplot2* (version 3.3.5). The results of GO analysis were visualized by a web tool *Revigo*, which shows the cluster representatives in a two-dimensional space derived by applying multidimensional scaling to a matrix of the GO terms' semantic similarities (Supek et al., 2011).

In order to validate the function of SFs, violin plots visualized by *ggplot2* (version 3.3.5) were used for verifying the difference in the expression of SFs in tumor and normal tissues. The paired-samples *t* test was used to test the significance of the difference.

The Kaplan-Meier (KM), generated by *survival* (version 3.2-11) and *survminer* (version 0.4.9), was applied to validate the prognostic effect of SFs (**Supplementary Source Code S3**) (Dinse and Lagakos, 1982). The log-rank test was used to test the significance of differences in survival between high- and low-risk patients (Mantel, 1966). The *p*-value  $< 0.05$  was considered statistically significant in this study.

## RESULTS

### Clinical Characteristics of the Lung Squamous Cell Carcinoma Cohort

The current study included a total of 493 LUSC patients from the TCGA database, and the characteristics and clinical information of these patients are listed in **Table 1**. There were 365 men and 128 women among these patients. With a median age of 68 (ranging from 39 to 85 years old), the mean survival time of patients was 1,044 days (ranging from 4 to 4,765 days). It is worth noting that the survival time of patients is censored data. The patient mortality rate of 43% confirms that LUSC is a tumor with a high mortality rate. The LUSC tumor staging data show that most patients are in stages I or II. Stage I tumors are usually small, without lymph nodes and distant metastases, and can be completely removed by surgery. In contrast, higher stages mean that the tumor is more progressive.

### Overview of Alternative Splicing Events in the Lung Squamous Cell Carcinoma Cohort

The TCGASpliceSeq database recorded seven types of ASEs, including exon skipping (ES), mutually exclusive (ME) exons, intron retention (RI), alternative promoter (AP), alternative

**TABLE 1** | Clinical characteristics of 493 LUSC patients in the TCGA database.

Characteristics	Groups	No. of patients	%
Sex	Male	365	74
	Female	128	26
Age at diagnosis	Median	68	
	Range	39–85	
	<61	107	22
	≥61	381	77
Vital status	Unknown	5	1
	Alive	268	54
	Dead	211	43
Stage	Unknown	14	3
	I	241	48.88
	II	158	32.05
	III	83	16.84
	IV	7	1.42
T category	Unknown	4	0.81
	T1	114	23
	T2	286	58
	T3	70	14
	T4	23	5
N category	N0	316	64
	N1	127	26
	N2	40	8
	N3	5	1
M category	NX	5	1
	M0	405	82
	M1	7	1
	MX	77	16
	Unknown	4	1

terminator (AT), alternative donor (AD) site, and alternative acceptor (AA) site (**Figure 2A**).

In this cohort, a total of 9424 ASEs were in 4246 genes, with 3876 ATs in 1817 genes, 2048 ESs in 1488 genes, 1761 APs in 791 genes, 656 RIs in 504 genes, 531 AAs in 468 genes, 515 ADs in 421 genes, and 37 MEs in 37 genes. Multiple ASEs can occur in a single gene (**Figure 2B**).

## Screening and Analysis of Alternative Splicing Events Related to Survival

Due to the complexity of biological processes, synergistic interactions between genes are more prevalent. To accurately screen OS-ASEs, we employed the *p*-value of the likelihood ratio test in bivariate Cox proportional risk regression and AUROC as screening criteria ( $P < 0.05$ ,  $AUROC > 0.8$ ). Consequently, 1118 combinations of OS-ASEs were screened, including 953 non-redundant ASEs.

A total of 953 OS-ASEs were detected in 489 genes. More specifically, there were 689 ATs in 348 genes, 241 APs in 121 genes, 12 ESs in 12 genes, 10 RIs in 10 genes, and 1 AA in 1 gene. Two splice types, AD and ME, were not included (**Figure 3A**). Next, in order to understand the function of the genes corresponding to OS-ASEs, KEGG analysis and GO analysis were performed. KEGG analysis demonstrated that these genes were enriched in histone-lysine N-methyltransferase activity, protein tyrosine phosphatase activity, and DNA repair and apoptosis pathways. These pathways are closely associated

**TABLE 2** | Top 10 SFs for degree centrality.

Rank	Ensembl	Symbol
1	ENSG00000130332	LSM7
2	ENSG00000169976	SF3B5
3	ENSG00000163634	THOC7
4	ENSG00000051596	THOC3
5	ENSG00000108561	C1QBP
6	ENSG00000137168	PP1L1
7	ENSG00000079134	THOC1
8	ENSG00000139343	SNRPF
9	ENSG00000108883	EFTUD2
10	ENSG00000123154	WDR83

with cancer progression (Östman et al., 2006; Wong, 2011; Jeggo et al., 2016; Husmann and Gozani, 2019). A recent study has shown that histone-lysine N-methyltransferase is a key driver for the induction of LUSC (**Figure 3B**) (Yuan et al., 2021). GO analysis revealed that these genes were enriched in both the nucleus and cytoplasm and play a role in protein binding, nucleic acid binding, and histone lysine N-methyltransferase activity. These genes are involved in important biological processes such as DNA repair, peptidyltyrosine dephosphorylation, and apoptosis (**Figure 3C**).

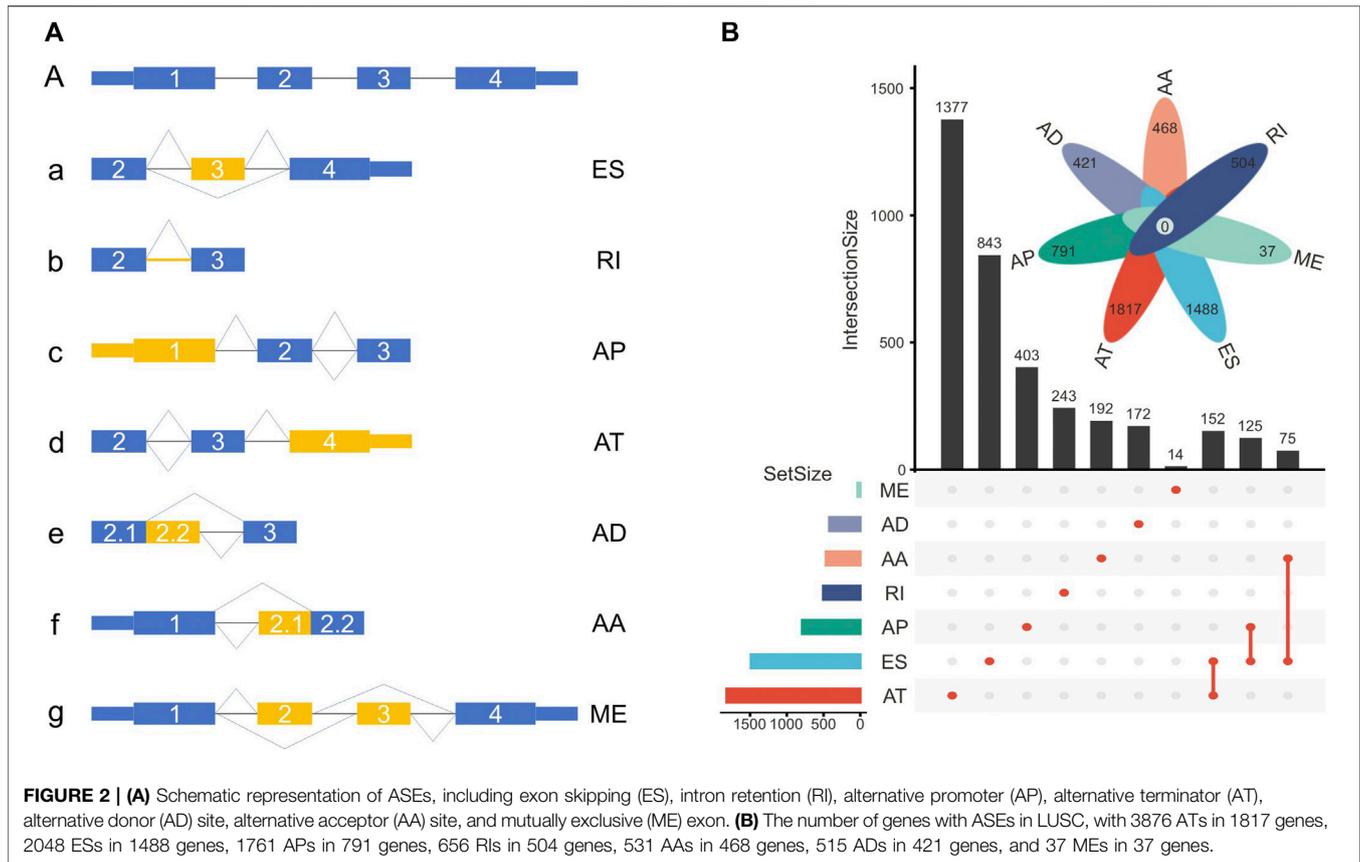
## Construction and Analysis of the Association Network Between Splicing Factors and Alternative Splicing Events

Systems biology is the study of the composition and interrelationships of the biological systems and is widely used in the study of gene networks (Hood, 2003). For our association network, identifying key vertices is an important way to find key SFs (Zhao and Liu, 2019). The association network was formed with 489 ASEs and 398 SFs as vertices and 9414 pairs of significant correlations as edges (**Figure 4**). The degree distribution is shown in **Supplementary Figure S1**. The average degree of the top 10 vertices in this network is 69, and the average degree of the remaining vertices is 22, indicating that the top 10 SFs are associated with more ASEs and is important in this network. Therefore, we consider these 10 SFs as key SFs. (**Table 2**)

## Validation of Splicing Factors

To verify the validity of the above approach, we analyzed the expression patterns of the 10 SFs in the TCGA-LUSC dataset. It is noticed that a significant difference exists in the expression of the 10 SFs between cancerous and paracancerous tissues (**Figure 5**). Moreover, patients were divided into two groups according to the expression of SFs, and the difference of survival time between them was analyzed with KM curves. It is found that 5 of these 10 SFs are significantly associated with the prognosis of LUSC patients. (**Supplementary Figure S2**).

In order to further assess the applicability of our approach, three cross-platform datasets from the GEO dataset were recruited. The GSE157010 dataset matches 9 SFs, 6 of which exhibit prognostic function (**Supplementary Figure S3**). In the



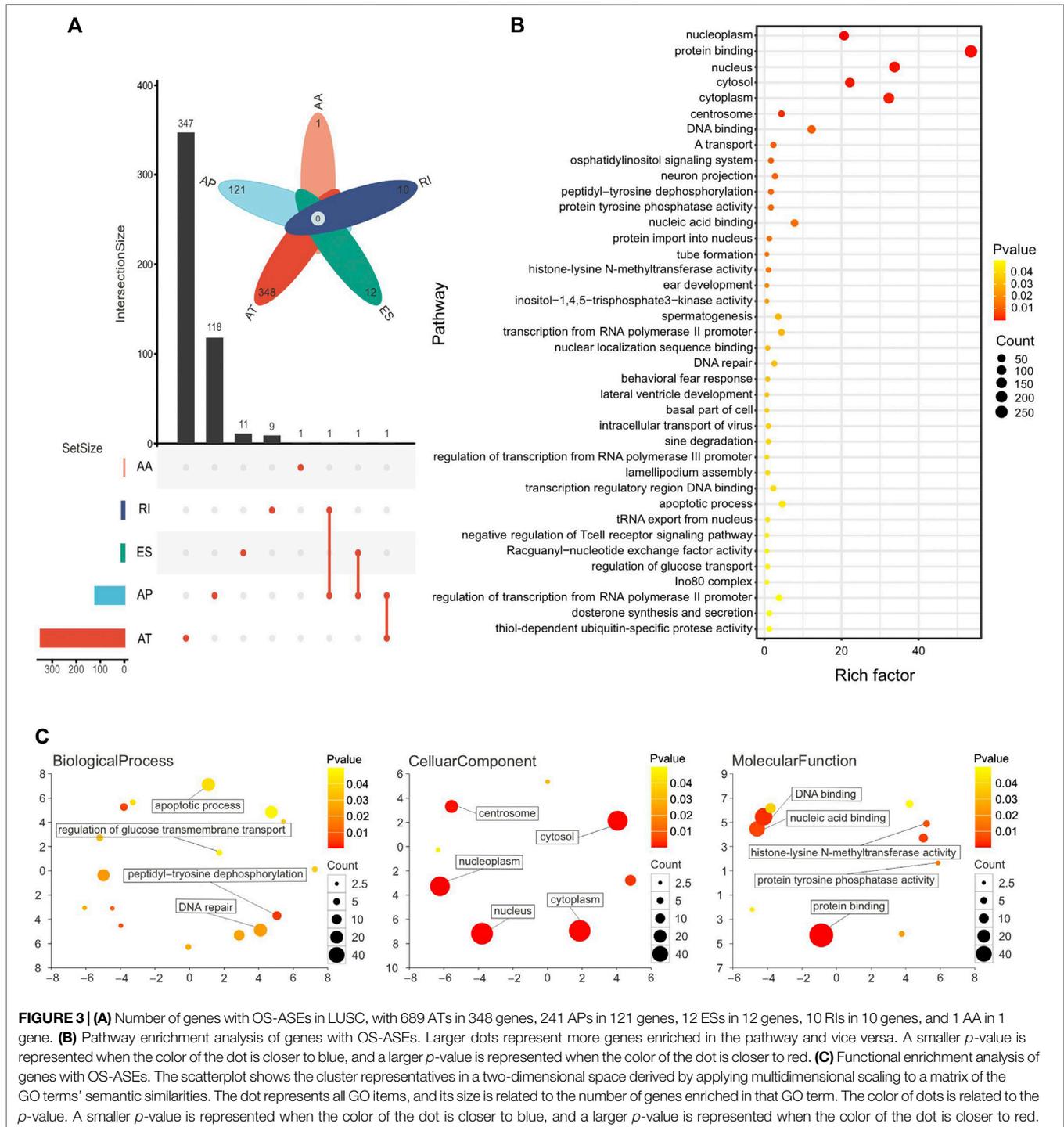
GSE3268 and GSE6044 datasets, the matched SFs are differentially expressed in normal and tumor samples (**Supplementary Figure S4**). In the GSE6044 dataset, the expression levels of SFs patients who received platinum-based therapy are slightly decreased compared with that of patients who did not, which is closer to the expression level in normal tissues (**Supplementary Figure S5**).

## DISCUSSION

In this research, bivariate Cox regression and the systems biology approach were employed to detect OS-ASEs and SFs associated with LUSC. The results showed that all 10 candidates (SFs) were expressed at significantly higher levels in tumor samples than in paracancerous tissues in both the TCGA-LUSC and GEO datasets. Moreover, 7 of these SFs were associated with the overall survival time in tumor patients in one or more datasets. These results are consistent with the currently known characteristics of tumor-associated genes (Givechian et al., 2018; Qi et al., 2018). It is found that 3 of the 10 SFs are reported to be connected with lung cancer, namely LSM7, C1QBP, and THOC1. Specifically, LSM7 is a prognosis-related key gene and mediates autophagy in LUSC, with significant expression differences between tumor and normal tissues (Gatica et al., 2019; Li et al., 2020); C1QBP is involved in various cellular processes,

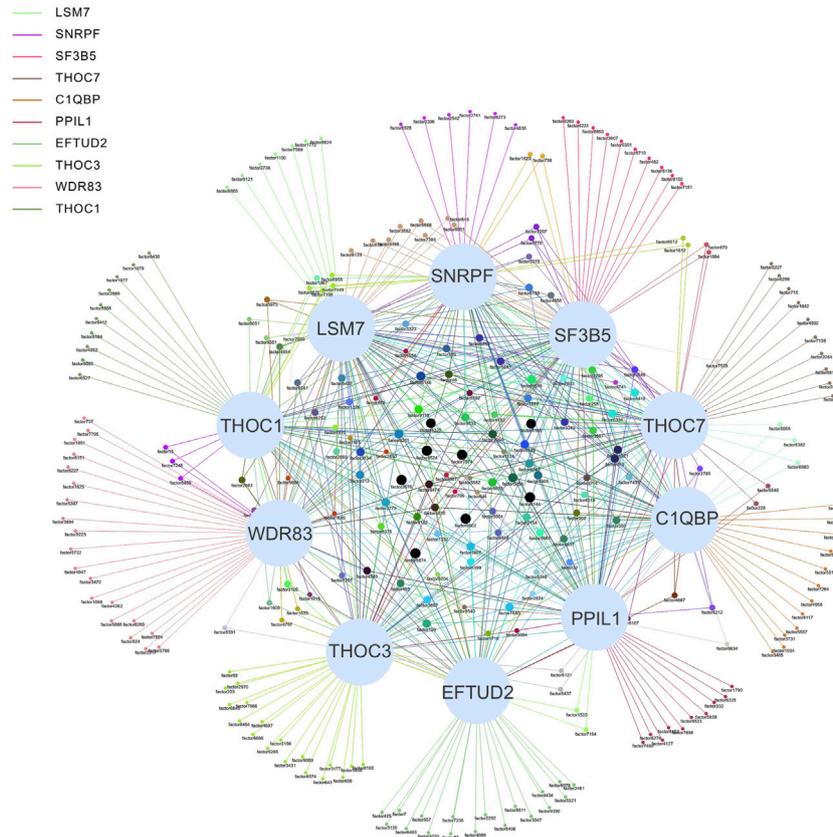
including mRNA splicing, ribosome biosynthesis, protein synthesis in mitochondria, apoptosis, transcriptional regulation, and viral infection, and its expression correlated with the prognosis of patients with lung, breast, and colon tumors (Saha et al., 2019); THOC1 is down-regulated in lung cancer cell lines SPC-A1 and NCI-H1975, and its overexpression inhibits cell proliferation, induces G2/M cell cycle arrest, and promotes cell apoptosis (Wan et al., 2014). THOC1 also inhibits the proliferation of tumor cells in hepatocellular carcinoma and prostate cancer (Liu et al., 2015; Cai et al., 2020). The above evidence suggests that our method is reliable and accurate.

In addition, we identified 7 new SFs, 6 of which, including SF3B5, THOC7, THOC3, SNRPF, EFTUD2, and WDR83, were reported to be associated with other tumors. It has been suggested that SF3B5 is a key prognostic factor in ovarian cancer (Ouyang et al., 2021). Studies have shown a relationship between the downregulation of THOC7 and the activation of tumorigenic pathways in cervical cancer (Lando et al., 2013; Lando et al., 2015). THOC3 is involved in the THO subcomplex and is necessary for coupled mRNA transcriptional extension and nuclear export, and its expression is significantly elevated in glioma cells (Chen Z. et al., 2021). SNRPF is aberrantly expressed in human glioma. *In vitro* experiments have revealed that ubiquitin carboxy-terminal hydrolase isozyme L5 could inhibit human glioma cell migration and invasion by downregulating SNRPF (Ge et al., 2017). EFTUD2 is markedly

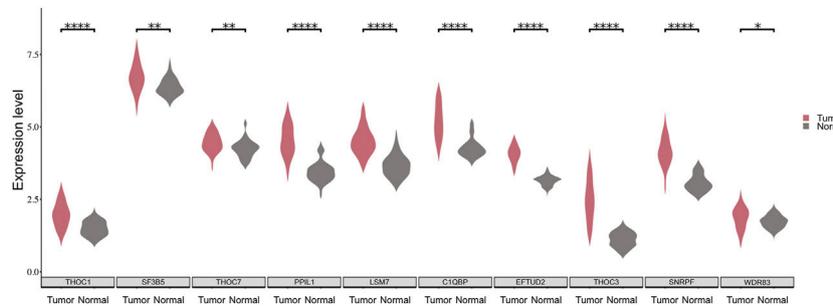


overexpressed in hepatocellular carcinoma tissues. High expression of EFTUD2 in hepatocellular carcinoma patients is associated with clinical features and is pivotal in hepatocellular carcinoma cell proliferation and cell cycle count (Lv et al., 2021). As the NAT of WDR83, the protein-coding gene, deoxyhypusine synthase, concordantly regulates the expressions of WDR83 mRNA and protein. Conversely, WDR83 also regulates

deoxyhypusine synthase by antisense pairing concordantly. As a pair of protein-coding cis-sense/antisense transcripts, WDR83 and DHPS are upregulated simultaneously and correlate positively in lung cancer. They drive the pathophysiology of lung cancer by promoting cell proliferation (Su et al., 2012). Furthermore, the remaining SF PP1L1, which has not been directly reported in the literature to be associated with cancer,



**FIGURE 4 |** Network diagram of the top 10 SFs and OS-ASEs. The 10 large light blue dots represent 10 SFs, the small dots represent ASEs, and the edges represent significant correlations between the two dots. Edges of different colors represent associations with different SFs. When an ASE is associated with multiple SFs, the color of the edge is a superposition of the corresponding multiple colors. The correspondence between the factor and ASE is shown in **Supplementary Table S3**.



**FIGURE 5 |** The expression distribution of 10 SFs between cancerous and paraneoplastic tissues in the TCGA dataset. The horizontal axis represents the expression of genes. The vertical axis shows the 10 key SFs.

is a member of the peptidyl-prolyl isomerase procyclin family and is frequently overexpressed in colon cancer cells (Chai et al., 2021). In summary, it is reasonable to speculate that the 7 SFs may play a role in the development of tumors, and the relationship between these SFs and lung cancer warrants further exploration in the future.

Our analysis method can be used not only to screen for key SFs in LUSC but also to apply to a wider range of studies. From the perspective of the study object, although our method is only

applied to LUSC data in this study, it is also applicable to other tumor data. From the perspective of research objectives, our method is not limited to screening SFs, but also can be used to screen regulatory factors, such as transcription factors, miRNAs or lncRNAs. For example, we can screen combinations of genes that can accurately classify tumor samples by downscaling or regression and then find key vertices by constructing a regulatory network of miRNAs that can anchor key miRNAs associated with tumors.

In conclusion, our analytical approach with a wide range of applications helps to obtain proper results and can provide new directions and perspectives for the exploration of related studies. In our study, although the specific functions and mechanisms of the 10 key SFs need to be further investigated, the available data and literature imply that they play a critical role in LUSC. Seven of these new SFs are also expected to be a new focus for future studies on SFs in LUSC. Furthermore, our proposed method will provide ideas and references for more studies. However, some limitations remain in our study. Due to the complexity of calculating multivariate combinations, we only calculated bivariate combinations, but multifactor combinations were not further explored. In subsequent studies, we will further improve our methods and extend to more scientific questions to provide novel focuses for future research.

## CONCLUSION

Abnormal AS is widely considered a novel indicator of carcinogenic processes, and SFs play a vital role in this process. Consequently, our aim is to screen key SFs that regulate carcinogenesis and progression. All combinations consisting of two ASEs were first constructed and screened using bivariate Cox regression and AUROC. Next, an association network of OS-ASEs and SFs was constructed by the Spearman correlation. Based on topological properties, we screened the top 10 SFs in terms of degree centrality. Finally, literature and data validation were performed on these 10 SFs. The data validation showed that 10 SFs were all significantly differentially expressed in both cancerous and paracancerous tissues of LUSC patients. Moreover, 5 of these SFs showed prognostic effects. It has been reported that 8 of these SFs are

## REFERENCES

- Anczuków, O., and Krainer, A. R. (2016). Splicing-factor Alterations in Cancers. *RNA* 22, 1285–1301. doi:10.1261/rna.057919.116
- Baralle, F. E., and Giudice, J. (2017). Alternative Splicing as a Regulator of Development and Tissue Identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi:10.1038/nrm.2017.27
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for Functional Genomics Data Sets-Update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193
- Bishara, A. J., and Hittner, J. B. (2012). Testing the Significance of a Correlation with Nonnormal Data: Comparison of Pearson, Spearman, Transformation, and Resampling Approaches. *Psychol. Methods* 17, 399–417. doi:10.1037/a0028087
- Bradburn, M. J., Clark, T. G., Love, S. B., and Altman, D. G. (2003). Survival Analysis Part II: Multivariate Data Analysis - an Introduction to Concepts and Methods. *Br. J. Cancer* 89, 431–436. doi:10.1038/sj.bjc.6601119
- Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments. *BMC Bioinformatics* 11, 94. doi:10.1186/1471-2105-11-94
- Cai, S., Bai, Y., Wang, H., Zhao, Z., Ding, X., Zhang, H., et al. (2020). Knockdown of THOC1 Reduces the Proliferation of Hepatocellular Carcinoma and Increases the Sensitivity to Cisplatin. *J. Exp. Clin. Cancer Res.* 39, 135. doi:10.1186/s13046-020-01634-7

closely associated with tumors. In addition, cross-platform validations of GEO were carried out, and similar results were obtained. These findings can serve as a reference for subsequent experimental studies.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA database (<https://portal.gdc.cancer.gov/projects/TCGA-LUSC>), TCGASpliceSeq database (<https://bioinformatics.mdanderson.org/TCGASpliceSeq/index.jsp>).

## AUTHOR CONTRIBUTIONS

MC and RZ contributed to the data acquisition, figure drawing and drafting of the manuscript. LZ and FZ contributed to the data analysis and interpretation, conception and design of the study. All authors read and approved the final manuscript.

## FUNDING

This work was supported by grants from National Natural Science Foundation of China (Grant No. 32070395).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.803606/full#supplementary-material>

- Chai, G., Webb, A., Li, C., Antaki, D., Lee, S., Breuss, M. W., et al. (2021). Mutations in Spliceosomal Genes PP1L1 and PRP17 Cause Neurodegenerative Pontocerebellar Hypoplasia with Microcephaly. *Neuron* 109, 241–256. doi:10.1016/j.neuron.2020.10.035
- Chen, L., Yao, Y., Sun, L., Zhou, J., Miao, M., Luo, S., et al. (2017). Snail Driving Alternative Splicing of CD44 by ESRP1 Enhances Invasion and Migration in Epithelial Ovarian Cancer. *Cell. Physiol. Biochem.* 43, 2489–2504. doi:10.1159/000484458
- Chen, T., Zhang, H., Liu, Y., Liu, Y.-X., and Huang, L. (2021a). EVenN: Easy to Create Repeatable and Editable Venn Diagrams and Venn Networks Online. *J. Genet. Genomics* 48, 863–866. doi:10.1016/j.jgg.2021.07.007
- Chen, Z., Wu, H., Yang, H., Fan, Y., Zhao, S., and Zhang, M. (2021b). Identification and Validation of RNA-binding Protein-related Gene Signature Revealed Potential Associations with Immunosuppression and Drug Sensitivity in Glioma. *Cancer Med.* 10, 7418–7439. doi:10.1002/cam4.4248
- Cheng, Z., Yu, C., Cui, S., Wang, H., Jin, H., Wang, C., et al. (2019). circTP63 Functions as a ceRNA to Promote Lung Squamous Cell Carcinoma Progression by Upregulating FOXM1. *Nat. Commun.* 10, 3200. doi:10.1038/s41467-019-11162-4
- Climente-González, H., Porta-Pardo, E., Godzik, A., and Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. *Cel Rep.* 20, 2215–2226. doi:10.1016/j.celrep.2017.08.012
- David, C. J., and Manley, J. L. (2010). Alternative Pre-mRNA Splicing Regulation in Cancer: Pathways and Programs Unhinged. *Genes Dev.* 24, 2343–2364. doi:10.1101/gad.1973010

- Dinse, G. E., and Lagakos, S. W. (1982). Nonparametric Estimation of Lifetime and Disease Onset Distributions from Incomplete Observations. *Biometrics* 38, 921. doi:10.2307/2529872
- Dorman, S. N., Viner, C., and Rogan, P. K. (2014). Splicing Mutation Analysis Reveals Previously Unrecognized Pathways in Lymph Node-Invasive Breast Cancer. *Sci. Rep.* 4. doi:10.1038/srep07063
- Dvinge, H., Kim, E., Abdel-Wahab, O., and Bradley, R. K. (2016). RNA Splicing Factors as Oncoproteins and Tumour Suppressors. *Nat. Rev. Cancer* 16, 413–430. doi:10.1038/nrc.2016.51
- Freeman, L. C. (1978). Centrality in Social Networks Conceptual Clarification. *Social Networks* 1, 215–239. doi:10.1016/0378-8733(78)90021-7
- Gatica, D., Hu, G., Liu, X., Zhang, N., Williamson, P. R., and Klionsky, D. J. (2019). The Pat1-Lsm Complex Stabilizes ATG mRNA during Nitrogen Starvation-Induced Autophagy. *Mol. Cell* 73, 314–324. doi:10.1016/j.molcel.2018.11.002
- Ge, J., Hu, W., Zhou, H., Yu, J., Sun, C., and Chen, W. (2017). Ubiquitin Carboxyl-Terminal Hydrolase Isozyme L5 Inhibits Human Glioma Cell Migration and Invasion via Downregulating SNRPF. *Oncotarget* 8, 113635–113649. doi:10.18632/oncotarget.23071
- Givechian, K. B., Wnuk, K., Garner, C., Benz, S., Garban, H., Rabizadeh, S., et al. (2018). Identification of an Immune Gene Expression Signature Associated with Favorable Clinical Features in Treg-Enriched Patient Tumor Samples. *Npj Genomic Med.* 3, 1–12. doi:10.1038/s41525-018-0054-7
- Han, J., Li, J., Ho, J. C., Chia, G. S., Kato, H., Jha, S., et al. (2017). Hypoxia Is a Key Driver of Alternative Splicing in Human Breast Cancer Cells. *Sci. Rep.* 7, 4108. doi:10.1038/s41598-017-04333-0
- Hazra, A., and Gogtay, N. (2017). Biostatistics Series Module 9: Survival Analysis. *Indian J. Dermatol.* 62, 251–257. doi:10.4103/ijid.IJD\_201\_17
- Hood, L. (2003). Systems Biology: Integrating Technology, Biology, and Computation. *Mech. Ageing Dev.* 124, 9–16. doi:10.1016/S0047-6374(02)00164-1
- Hu, Y.-X., Zheng, M.-J., Zhang, W.-C., Li, X., Gou, R., Nie, X., et al. (2019). Systematic Profiling of Alternative Splicing Signature Reveals Prognostic Predictor for Cervical Cancer. *J. Transl. Med.* 17, 379. doi:10.1186/s12967-019-02140-x
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4, 44–57. doi:10.1038/nprot.2008.211
- Husmann, D., and Gozani, O. (2019). Histone Lysine Methyltransferases in Biology and Disease. *Nat. Struct. Mol. Biol.* 26, 880–889. doi:10.1038/s41594-019-0298-7
- Jeggo, P. A., Pearl, L. H., and Carr, A. M. (2016). DNA Repair, Genome Stability and Cancer: A Historical Perspective. *Nat. Rev. Cancer* 16, 35–42. doi:10.1038/nrc.2015.4
- Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., et al. (2010). Identification of Influential Spreaders in Complex Networks. *Nat. Phys.* 6, 888–893. doi:10.1038/nphys1746
- Lando, M., Fjeldbo, C. S., Wilting, S. M., Snoek, B. C., Aarnes, E.-K., Forsberg, M. F., et al. (2015). Interplay between Promoter Methylation and Chromosomal Loss in Gene Silencing at 3p11-P14 in Cervical Cancer. *Epigenetics* 10, 970–980. doi:10.1080/15592294.2015.1085140
- Lando, M., Wilting, S. M., Snipstad, K., Clancy, T., Bierkens, M., Aarnes, E.-K., et al. (2013). Identification of Eight Candidate Target Genes of the Recurrent 3p12-P14 Loss in Cervical Cancer by Integrative Genomic Profiling. *J. Pathol.* 230, 59–69. doi:10.1002/path.4168
- Li, W., Li, X., Gao, L.-N., and You, C.-G. (2020). Integrated Analysis of the Functions and Prognostic Values of RNA Binding Proteins in Lung Squamous Cell Carcinoma. *Front. Genet.* 11. doi:10.3389/fgene.2020.00185
- Li, Y., Gu, J., Xu, F., Zhu, Q., Ge, D., and Lu, C. (2018). Transcriptomic and Functional Network Features of Lung Squamous Cell Carcinoma through Integrative Analysis of GEO and TCGA Data. *Sci. Rep.* 8, 1–12. doi:10.1038/s41598-018-34160-w
- Linden, A. (2006). Measuring Diagnostic and Predictive Accuracy in Disease Management: An Introduction to Receiver Operating Characteristic (ROC) Analysis. *J. Eval. Clin. Pract.* 12, 132–139. doi:10.1111/j.1365-2753.2005.00598.x
- Liu, C., Yue, B., Yuan, C., Zhao, S., Fang, C., Yu, Y., et al. (2015). Elevated Expression of Thoc1 Is Associated with Aggressive Phenotype and Poor Prognosis in Colorectal Cancer. *Biochem. Biophysical Res. Commun.* 468, 53–58. doi:10.1016/j.bbrc.2015.10.166
- Lv, C., Li, X. J., Hao, L. X., Zhang, S., Song, Z., Ji, X. D., et al. (2021). Over-activation of EFTUD2 Correlates with Tumor Propagation and Poor Survival Outcomes in Hepatocellular Carcinoma. *Clin. Transl. Oncol.* doi:10.1007/s12094-021-02673-y
- Mantel, N. (1966). Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration. *Cancer Chemother. Rep.* 50, 163–170. Available at : <https://pubmed.ncbi.nlm.nih.gov/5910392/> (Accessed September 27, 2021).
- Matsuda, T., Namura, A., and Oinuma, I. (2019). Dynamic Spatiotemporal Patterns of Alternative Splicing of an F-Actin Scaffold Protein, Afadin, during Murine Development. *Gene* 689, 56–68. doi:10.1016/j.gene.2018.12.020
- Miura, K., Fujibuchi, W., and Unno, M. (2012). Splice Isoforms as Therapeutic Targets for Colorectal Cancer. *Carcinogenesis* 33, 2311–2319. doi:10.1093/carcin/bgs347
- Moncada, R., Barkley, D., Wagner, F., Chiodin, M., Devlin, J. C., Baron, M., et al. (2020). Integrating Microarray-Based Spatial Transcriptomics and Single-Cell RNA-Seq Reveals Tissue Architecture in Pancreatic Ductal Adenocarcinomas. *Nat. Biotechnol.* 38, 333–342. doi:10.1038/s41587-019-0392-8
- Nakka, K., Ghigna, C., Gabellini, D., and Dilworth, F. J. (2018). Diversification of the Muscle Proteome through Alternative Splicing. *Skeletal Muscle* 8, 8. doi:10.1186/s13395-018-0152-3
- Östman, A., Hellberg, C., and Böhmer, F. D. (2006). Protein-tyrosine Phosphatases and Cancer. *Nat. Rev. Cancer* 6, 307–320. doi:10.1038/nrc1837
- Ouyang, Y., Xia, K., Yang, X., Zhang, S., Wang, L., Ren, S., et al. (2021). Alternative Splicing Acts as an Independent Prognosticator in Ovarian Carcinoma. *Sci. Rep.* 11, 10413. doi:10.1038/s41598-021-89778-0
- 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. doi:10.1038/ng.259
- Park, S., Brugiolo, M., Akerman, M., Das, S., Urbanski, L., Geier, A., et al. (2019). Differential Functions of Splicing Factors in Mammary Transformation and Breast Cancer Metastasis. *Cel Rep.* 29, 2672–2688. doi:10.1016/j.celrep.2019.10.110
- Pradella, D., Naro, C., Sette, C., and Ghigna, C. (2017). EMT and Stemness: Flexible Processes Tuned by Alternative Splicing in Development and Cancer Progression. *Mol. Cancer* 16. doi:10.1186/s12943-016-0579-2
- Qi, C., Chen, Y., Zhou, Y., Huang, X., Li, G., Zeng, J., et al. (2018). Delineating the Underlying Molecular Mechanisms and Key Genes Involved in Metastasis of Colorectal Cancer via Bioinformatics Analysis. *Oncol. Rep.* 39, 2297–2305. doi:10.3892/or.2018.6303
- Ryan, M., Wong, W. C., Brown, R., Akbani, R., Su, X., Broom, B., et al. (2016). TCGASpliceSeq a Compendium of Alternative mRNA Splicing in Cancer. *Nucleic Acids Res.* 44, D1018–D1022. doi:10.1093/nar/gkv1288
- Saha, S., Kim, K., Islam, S. M., Cho, S.-G., and Gil, M. (2019). Systematic Multiomics Analysis of Alterations in C1QBP mRNA Expression and Relevance for Clinical Outcomes in Cancers. *Jcm* 8, 513. doi:10.3390/jcm8040513
- Shkreta, L., Bell, B., Revil, T., Venables, J. P., Prinos, P., Elela, S. A., et al. (2013). Cancer-associated Perturbations in Alternative Pre-messenger RNA Splicing. *Cancer Treat. Res.* 158, 41–94. doi:10.1007/978-3-642-31659-3\_3
- Shkreta, L., and Chabot, B. (2015). The RNA Splicing Response to DNA Damage. *Biomolecules* 5, 2935–2977. doi:10.3390/biom5042935
- Su, H., Hu, J., Huang, L., Yang, Y., Thenoz, M., Kuchmiy, A., et al. (2018). SHQ1 Regulation of RNA Splicing Is Required for T-Lymphoblastic Leukemia Cell Survival. *Nat. Commun.* 9, 4281. doi:10.1038/s41467-018-06523-4
- Su, W.-Y., Li, J.-T., Cui, Y., Hong, J., Du, W., Wang, Y.-C., et al. (2012). Bidirectional Regulation between WDR83 and its Natural Antisense Transcript DHPS in Gastric Cancer. *Cell Res* 22, 1374–1389. doi:10.1038/cr.2012.57
- Suo, C., Hrydziuszko, O., Lee, D., Pramana, S., Saputra, D., Joshi, H., et al. (2015). Integration of Somatic Mutation, Expression and Functional Data Reveals Potential Driver Genes Predictive of Breast Cancer Survival. *Bioinformatics* 31, 2607–2613. doi:10.1093/bioinformatics/btv164
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* 6, e21800. doi:10.1371/journal.pone.0021800

- Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A., and Skotheim, R. I. (2016). Aberrant RNA Splicing in Cancer; Expression Changes and Driver Mutations of Splicing Factor Genes. *Oncogene* 35, 2413–2427. doi:10.1038/onc.2015.318
- Tyson-Capper, A., and Gautrey, H. (2018). Regulation of Mcl-1 Alternative Splicing by hnRNP F, H1 and K in Breast Cancer Cells. *RNA Biol.* 15, 1448–1457. doi:10.1080/15476286.2018.1551692
- Wan, J., Zou, S., Hu, M., Zhu, R., Xu, J., Jiao, Y., et al. (2014). Thoc1 Inhibits Cell Growth via Induction of Cell Cycle Arrest and Apoptosis in Lung Cancer Cells. *Mol. Med. Rep.* 9, 2321–2327. doi:10.3892/mmr.2014.2088
- Wang, C., Tan, S., Liu, W.-R., Lei, Q., Qiao, W., Wu, Y., et al. (2019). RNA-seq Profiling of Circular RNA in Human Lung Adenocarcinoma and Squamous Cell Carcinoma. *Mol. Cancer* 18, 134. doi:10.1186/s12943-019-1061-8
- Wang, F., Fu, X., Chen, P., Wu, P., Fan, X., Li, N., et al. (2017). SPSB1-mediated HnRNP A1 Ubiquitylation Regulates Alternative Splicing and Cell Migration in EGF Signaling. *Cel Res* 27, 540–558. doi:10.1038/cr.2017.7
- Wang, Y., Yang, F., Shang, J., He, H., and Yang, Q. (2021). Integrative Analysis Reveals the Prognostic Value and Functions of Splicing Factors Implicated in Hepatocellular Carcinoma. *Sci. Rep.* 11, 1–14. doi:10.1038/s41598-021-94701-8
- Wong, A. C. H., Rasko, J. E. J., and Wong, J. J.-L. (2018). We Skip to Work: Alternative Splicing in normal and Malignant Myelopoiesis. *Leukemia* 32, 1081–1093. doi:10.1038/s41375-018-0021-4
- Wong, R. S. (2011). Apoptosis in Cancer: From Pathogenesis to Treatment. *J. Exp. Clin. Cancer Res.* 30, 87. doi:10.1186/1756-9966-30-87
- Wu, Z., Chen, H., Liang, Y., Luo, W., Deng, F., and Zeng, F. (2020). Alternative Splicing Implicated in Immunity and Prognosis of colon Adenocarcinoma. *Int. Immunopharmacology* 89, 107075. doi:10.1016/j.intimp.2020.107075
- Xie, R., Chen, X., Chen, Z., Huang, M., Dong, W., Gu, P., et al. (2019). Polypyrimidine Tract Binding Protein 1 Promotes Lymphatic Metastasis and Proliferation of Bladder Cancer via Alternative Splicing of MEIS2 and PKM. *Cancer Lett.* 449, 31–44. doi:10.1016/j.canlet.2019.01.041
- Xu, Q., Modrek, B., and Lee, C. (2002). Genome-wide Detection of Tissue-specific Alternative Splicing in the Human Transcriptome. *Nucleic Acids Res.* 30, 3754–3766. doi:10.1093/nar/gkf492
- Yuan, G., Flores, N. M., Hausmann, S., Lofgren, S. M., Kharchenko, V., Angulo-Ibanez, M., et al. (2021). Elevated NSD3 Histone Methylation Activity Drives Squamous Cell Lung Cancer. *Nature* 590, 504–508. doi:10.1038/s41586-020-03170-y
- Yuan, J.-h., Liu, X.-n., Wang, T.-t., Pan, W., Tao, Q.-f., Zhou, W.-p., et al. (2017). The MBNL3 Splicing Factor Promotes Hepatocellular Carcinoma by Increasing PXN Expression through the Alternative Splicing of lncRNA-PXN-AS1. *Nat. Cel Biol.* 19, 820–832. doi:10.1038/ncb3538
- Zhao, J., Chang, L., Gu, X., Liu, J., Sun, B., and Wei, X. (2020). Systematic Profiling of Alternative Splicing Signature Reveals Prognostic Predictor for Prostate Cancer. *Cancer Sci.* 111, 3020–3031. doi:10.1111/cas.14525
- Zhao, X., and Liu, Z.-P. (2019). Analysis of Topological Parameters of Complex Disease Genes Reveals the Importance of Location in a Biomolecular Network. *Genes* 10, 143. doi:10.3390/genes10020143
- Zhu, J., Chen, Z., and Yong, L. (2018). Systematic Profiling of Alternative Splicing Signature Reveals Prognostic Predictor for Ovarian Cancer. *Gynecol. Oncol.* 148, 368–374. doi:10.1016/j.ygyno.2017.11.028
- Zong, Z., Li, H., Yi, C., Ying, H., Zhu, Z., and Wang, H. (2018). Genome-wide Profiling of Prognostic Alternative Splicing Signature in Colorectal Cancer. *Front. Oncol.* 8, 537. doi:10.3389/fonc.2018.00537

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Zhu, Zhang and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.