Identification of Prognostic Alternative Splicing Signature in Triple-negative breast cancer

CURRENT STATUS: UNDER REVIEW

Shuqi Zhou
China Medical University

Shu Guan
China Medical University

Xuemei Lv
China Medical University

Rong Ma
China Medical University

Yifan Zhang
China Medical University

Yuanyuan Yan
China Medical University

Jue Wang
China Medical University

Xiaoyu Sun
China Medical University

Mengcong Gao
China Medical University

Di Di
China Medical University

Hongliang Zhao
China Medical University

Li Han
China Medical University
Minjie Wei  
China Medical University  
weiminjiecmu@163.com  
Corresponding Author  
ORCiD: https://orcid.org/0000-0002-0782-6556

Miao He  
China Medical University

DOI:  
10.21203/rs.3.rs-17500/v1

SUBJECT AREAS  
Cancer Biology

KEYWORDS  
Alternative splicing, Triple-negative breast cancer, Prognosis, Splicing factor
Abstract

Background: Alternative splicing (AS) is a pervasive and vital mechanism involved in the progression of various cancer. Studies confirm the importance of prognostic value of AS events in tumor patients, but systematic analysis of AS in triple-negative breast cancer (TNBC) is still lacking.

Methods: Information from 115 TNBC patients from the Cancer Genome Atlas (TCGA) database were extracted. And we performed a comprehensive analysis of whole-genome AS events with corresponding clinical information to evaluate the roles of seven AS patterns. Prognostic analyses were performed with predictive models and splicing network built for TNBC patients.

Results: Among 28,744 mRNA AS events in 20,353 genes, we detected 1,428 AS in 138 important survival genes related to the overall survival of TNBC patients event. Through functional and pathway enrichment analysis, we found that these genes are involved in ubiquitin-mediated proteolytic pathways. At 1800 days overall survival, the area under the ROC curve for prognostic signatures was 0.8. It shows that this model is very effective in differentiating patient prognosis. The use of Spearman's test to establish a potential regulatory network between survival-related AS events and abnormal SF indicates a clear trend in the role of SF in TNBC.

Conclusions: In summary, we established a reliable and powerful TNBC prognostic signature. A splicing network that could be its underlying mechanism was discovered.

Keywords: Alternative splicing; Triple-negative breast cancer; Prognosis; Splicing factor

Background

Breast cancer (BC) is the most commonly diagnosed cancer in women. By 2020, an estimated 276,480 women will develop breast cancer in the United States, accounting for 30 percent of the 912,930 new cancer cases diagnosed in women. BC is also the second leading reason of cancer deaths in women (15%) (Siegel, Miller, & Jemal, 2020). Clearly, breast cancer has been one of the greatest threats to women's health. Triple-negative breast cancer (TNBC) is a high-grade tumor with poor prognosis, accounting for 15–20% of BC, which characterized by lacking expression of the estrogen receptor (ER), progesterone receptor (PgR) and human epidermal growth factor receptor 2 (HER2) (Crisciitiello et al., 2018). However, compared to other molecular subtypes of BC, TNBC lacks a
clear molecular target. At present, chemotherapy is still one of the more commonly used treatments (de Nonneville et al., 2019). Accordingly, further screening of potential diagnostic biomarkers to treat TNBC is imminent.

Alternative splicing (AS) is a pervasive and vital mechanism involved in the progression of various cancer by expanding genomic encoding capacity and increasing protein complexity (Zhang, Duan, Wang, & Lin, 2019). AS was involved in a series of oncogenic processes, such as proliferation, metastasis, apoptosis and immune escape (David & Manley, 2010). There have been more evidences prove that AS is exploited by tumor cells, cancer cells depend on specific isoforms at the stage of carcinogenesis and disease development (Genov, Basti, Abreu, Astaburuaga, & Relogio, 2019). And under certain circumstances, the mutations of splicing factors (SFs) can produce specific cancer-promoting splicing isoforms (Zhu, Chen, & Yong, 2018). The prevalence of AS in tumors has been reported, and AS was associated with tumor occurrence, progression and treatment (Lee & Abdel-Wahab, 2016; Lu et al., 2015; Trincado, Sebestyen, Pages, & Eyras, 2016). Although there is growing report that the prognostic value of AS events in colorectal, papillary thyroid, and lung cancer patients (Li et al., 2017; Lin et al., 2019; Zong et al., 2018). However, systematic analysis of AS in TNBC has not been reported. Therefore, finding the potential regulatory relationship between AS events and SFs in TNBC is very necessary.

In conclusion, our study made the first attempt to integrate splicing profiles and TCGA clinical factors of TNBC patients to comprehensively research the prognostic value of AS and constructed prognostic signature for TNBC patients. Furthermore, in addition to valuable prognostic factors for patients, the present study uncovered interesting splicing networks in TNBC further revealing the potential mechanism of AS in TNBC tumorigenesis. Our results unravel the pattern of global aberrant AS and its clinical implications in TNBC and provide the evidence for further screening and identifying the potential biomarkers for the early diagnosis of TNBC.

Methods
Data collection and processing
Download the RNA sequencing data from the BRCA queue and the corresponding clinical information
from the TCGA data portal. After removing patients with incomplete information, 122 TNBC samples were obtained.

Meanwhile, we collected data of AS events using a Percent Spliced In (PSI) value of exceeding 75% in TNBC samples from the TCGA SpliceSep database to develop the AS profiling for every TNBC patient. The PSI value was an intuitive quantification used to quantify splicing event. The PSI values range from 0 to 100 (%), suggesting a shift in splicing events (Ryan, Cleland, Kim, Wong, & Weinstein, 2012). The AS events include seven types: Exon Skip (ES), Mutually Exclusive Exons (ME), Retained Intron (RI), Alternate Promoter (AP), Alternate Terminator (AT), Alternate Donor site (AD), and Alternate Acceptor site (AA). These seven types of schematic diagrams were shown in Figure.1. In the end, 115 patients were enrolled in our cohort.

**Survival analysis**

We performed Univariate Cox regression analysis to assess the relationship between AS events and overall survival (OS), then applied Multivariate COX regression to analyze independent prognostic predictors. In order to compare the efficiencies of prognostic factors, the survival Receiver Operating Characteristic Curve (ROC) package in R was used to estimate the time-dependent ROC curve with review data, and the Area Under Curve (AUC) of the ROC curves were generated for these models (Li et al., 2017). The model's ability to predict the outcomes at the points in time were compared on account of fewer events occurred after 1800 days in all survival analyses. Finally, we constructed the risk score models according to a combination of gene expression levels weighted by regression coefficient ($\beta$) originating from the multivariate Cox regression analysis. The formula was as follows:

$$\text{Risk score} = \beta_{\text{gene1}} \times \text{expr}_{\text{gene1}} + \beta_{\text{gene2}} \times \text{expr}_{\text{gene2}} + \cdots + \beta_{\text{genen}} \times \text{expr}_{\text{genen}}.$$  

All analyses were performed using R software (version 3.4.1).

**Bioinformatics analyses**

Here, the UpSet plot for quantitative analysis of the interaction sets between seven sorts of AS was from UpSetR (version 1.3.3) (Lex, Gehlenborg, Strobelt, Vuillemot, & Pfister, 2014). Genetic network analysis was carried out by inputting the relevant gene identifiers of the survival-related AS genes into Cytoscape. Using it can search for important central genes in survival-associated AS genes. Use
Univariate Cox regression to analyze survival-associated SFs to further determine whether these SFs expression were significantly related to the PSI values of survival-related AS events. It was then generated through Cytoscape (version 3.7.1) to visualize the correlation network.

**Functional pathway enrichment analysis**

Using the annotations of DAVID (Version 6.8) to visualise and integrate database, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed for the genes with survival associated AS events (Sherman et al., 2007). GO and KEGG were used to assess the functional categories, terms with a P value and q-value both smaller than 0.05 were considered significant categories. P value < 0.05 was selected as the standard segmentation.

**Construction of Potential SFs-ASs Network**

A list of 404 SFs was got from a previous study (Supplementary Table 1) (Seiler et al., 2018). The expression profiles of SFs in the mRNA splicing pathways were extracted and standardized from the mRNA sequencing data in TCGA for further Univariate Cox regression analysis. Next, Spearman method was applied to analyze the correlation between the OS-associated SFs expression and the PSI values of survival-associated AS events included in the construction of each prognostic feature. By Benjamini and Hochberg (BH) correlation, P value < 0.05 was considered significant. Use Cytoscape (3.7.1) to generate a potential SFs-ASs regulatory network among the significantly correlated pairs.

**Results**

**The Overviews of mRNA splicing events in TNBC from TCGA**

A flowchart in Fig. 1 showed detailed processes of our study design. Integrated mRNA splicing events profiles for 115 TNBC patients were established in depth from RNA sequencing data. In Fig. 2A, seven sorts of AS events were illustrated. As a whole, there are 28,744 mRNA splicing events in 20,353 genes in TNBC cohort, which contains 3215 AA events in 2456 genes, 2624 AD events in 2101 genes, 3480 AP events in 3480 genes, 3715 AT events in 3715 genes, 12988 ES events in 6661 genes, 70 ME events in 70 genes, and 2652 RI events in 1870 genes for evaluating prognostic value (Fig. 2B).

**Survival associated-AS events in TNBC**

In order to investigate whether AS events were related to patient survival, we first analyzed the association between AS events and OS for the TNBC. By the univariate Cox regression, 1,428 AS events with a total of 1,383 genes were notably associated with OS in TNBC patients (P < 0.05,
Table 1). In detail, a total of 174 AA events in 169 genes, 143 AD events in 141 genes, 156 AP events in 156 genes, 170 AT events in 170 genes, 641 ES events in 606 genes, 2 ME events in 2 genes, and 142 RI events in 139 genes were identified as prognosis-associated AS events (P < 0.05, Fig. 2C). The UpSet plot vividly revealed that about half of the AS events were ES events, which was the most predominant event type related to prognosis. Interestingly, these results also markedly indicated that one gene could have as many as three mRNA splicing events associated with prognosis (Fig. 2D).

Table 2 listed the top 20 most relevant genes among the seven sorts of AS associated with patient survival. The most significant genes associated with survival in TNBC (P < 0.005) were introduced into cytoscape to generate a gene interaction network as shown in Fig. 3. Moreover, the gene network revealed important cancer pathways, and ANAPC7 (APC7), NEDD4L, UBE2N and UBE2F were central genes related to survival.

Table 1

| TNBCmarker | symbol    | splice_type |
|------------|-----------|-------------|
| ID 9457    | ZC3H11A   | AA          |
| ID 25959   | ATP7B     | ES          |
| ID 47225   | ANGPTL4   | RI          |
| ID 61125   | TANGO2    | ES          |
| ID 54993   | RGPD8     | AD          |
| ID 72360   | RAD17     | ES          |
| ID 66429   | CSTA      | AT          |
| ID 89523   | UPRT      | ES          |
| ID 80869   | ZKSCAN1   | ES          |
| ID 473     | PLEKHG5   | AA          |
| ID 69797   | FAM175A   | AP          |
| ID 88738   | GK        | ES          |
| ID 83788   | SPIDR     | ES          |
| ID 52113   | SSC5D     | AT          |
| ID 69930   | SNCA      | ES          |
| ID 37557   | ZFHX3     | AP          |
| ID 87859   | PPP2R4    | ES          |
| ID 1829    | EPHA10    | RI          |
| ID 1775    | OSCP1     | AP          |
| ID 64564   | MAP4      | ES          |
| ID 35582   | SCNN1B    | AA          |
| ID 26678   | HAUS4     | AD          |
| ID 18845   | NXPE2     | AT          |
| ID 85333   | SLC45A4   | RI          |
| ID 45655   | NEDD4L    | ES          |
| ID 33675   | CORO7     | ES          |
| ID 52592   | CMPK2     | AT          |
| ID 50311   | ZN1F33     | AT          |
| ID 51484   | ZNF83     | RI          |
| ID 13439   | FANK1     | AP          |
| ID 60850   | ITGB2     | ES          |
| ID 83027   | PDLIM2    | ES          |
| ID 16606   | PPP1R14B  | AP          |
| ID 40110   | CPD       | AP          |
| ID 79065   | TAX1BP1   | ES          |
| ID 74588   | CPEB4     | AP          |
| ID 84363   | WWP1      | ES          |
| ID   | Gene  | Abbrv |
|------|-------|-------|
| 45471 | DYM   | ES    |
| 72785 | FAM172A | ES   |
| 81379 | DLD   | AA    |
| 86783 | CDK20 | AA    |
| 12946 | NFKB2 | AP    |
| 49633 | CATSPERG | ES   |
| 11493 | WDFY4 | AT    |
| 69362 | SRP72 | ES    |
| 44009 | TBC1D16 | AP   |
| 71697 | AMACR | AT    |
| 35669 | LCMT1 | AD    |
| 94083 | TNRC6A | ES   |
| 36129 | ITGAL | ES    |
| 26387 | TFDP1 | AD    |
| 24901 | SED1B | AD    |
| 22026 | PRR13 | AP    |
| 29469 | KLC1  | ES    |
| 55935 | CYBRD1 | AP   |
| 509   | VAMP3 | AP    |
| 54642 | UNC50 | RI    |
| 69517 | RCHY1 | AA    |
| 25042 | PITPNM2 | AT   |
| 14339 | ADM   | RI    |
| 3207  | FGGY  | AP    |
| 10832 | SUV39H2 | AD   |
| 85269 | KHDRBS3 | AP  |
| 35513 | NPIPB4 | RI   |
| 30228 | LCMT2 | RI    |
| 44243 | PYCR1 | RI    |
| 19513 | SNX19 | AD    |
| 71011 | NPY1R | RI    |
| 35570 | NPIPB5 | RI   |
| 24085 | TGD   | ES    |
| 42874 | BCAS3 | ES    |
| 57100 | PARD3B | ES   |
| 25382 | ZNF268 | ES   |
| 38443 | TRPV3 | AT    |
| 56937 | SUMO1 | AA    |
| 24968 | RSRC2 | AA    |
| 63094 | SETD5 | AA    |
| 34285 | COQ7  | RI    |
| 31133 | PIF1  | AA    |
| 85994 | KLHL9 | RI    |
| 24422 | ANAPC7 | AT   |
| 53075 | PPP1CB | AD   |
| 76772 | COL12A1 | ES   |
| 83268 | TMEM66 | ES   |
| 65186 | ALAS1 | ES    |
| 26842 | DCAF11 | AD   |
| 85423 | MRONH6 | AP   |
| 33608 | DNASE1 | RI   |
| 46341 | MED16 | AA    |
| 31619 | STOML1 | AP   |
| 85835 | KDM4C | ES    |
| 67062 | ACPL2 | ES    |
| 37561 | PSMD7 | AA    |
| 42349 | ITGA3 | ES    |
| 21715 | TMPRSS12 | RI |
| 45739 | SERPINB8 | ES |
| 26673 | PRMT5 | AD    |
| 76012 | RNF8  | ES    |
| 42817 | TUBD1 | ES    |
| 80307 | DMTF1 | ES    |
| 66647 | GATA2 | AA    |
| 47313 | ZNF177 | ES   |
| 28068 | PLEKHH1 | AD |
| 27451 | VCPKMT | AD   |
| 85197 | PHF20L1 | AA |
| 83181 | CCDC25 | ES   |
| 84641 | RPL30 | RI    |
| ID_45697 | PIGN  | ES |
|----------|-------|----|
| ID 23688 | UBE2N | ES |
| ID 87156 | CTNNAL1 | ES |
| ID 11475 | MAPK8 | AA |
| ID 94142 | C16orf93 | ES |
| ID 82553 | ERICH1 | AT |
| ID 76336 | VEGFA | ES |
| ID 53776 | SLC1A4 | ES |
| ID 73374 | CDKL3 | ES |
| ID 17435 | FAM86C1 | AD |
| ID 58266 | DUSP28 | RI |
| ID 26303 | TUBGCP3 | AT |
| ID 86100 | NFX1 | AT |
| ID 85521 | PARP10 | RI |
| ID 25006 | OGFOD2 | AP |
| ID 87774 | SPTAN1 | ES |
| ID 13679 | RNH1 | AA |
| ID 15734 | SLC39A13 | AP |
| ID 62455 | MEI1 | ES |
| ID 69759 | LIN54 | ES |
| ID 23661 | DCN | ES |
| ID 90543 | HCFC1 | RI |
| ID 52702 | GREB1 | AT |
| ID 39047 | WRAP53 | RI |
| ID 64840 | USP19 | AA |
| ID 32441 | WDR93 | AT |
| ID 84857 | OXR1 | ES |
| ID 31201 | VWA9 | ES |
| ID 20354 | GABARAPL1 | AD |
| ID 18973 | AMICA1 | AD |
| ID 21837 | BIN2 | AP |
| ID 58475 | PSMF1 | AA |
| ID 42601 | SCPEP1 | ES |
| ID 29026 | UBR7 | ES |
| ID 59774 | TMEM189 | ES |
| ID 26652 | RBM23 | ES |
| ID 90655 | IKBKG | AA |
| ID 87307 | KIF12 | AA |
| ID 79802 | SUMF2 | ES |
| ID 58614 | PANK2 | AD |
| ID 64940 | RBM6 | ES |
| ID 78032 | SHPRH | AA |
| ID 31141 | ANKDD1A | ES |
| ID 58170 | UBE2F | ES |
| ID 65841 | ABI3BP | AT |
| ID 29010 | LGMN | ES |
| ID 53420 | SLC3A1 | AP |
| ID 42165 | NFE2L1 | RI |
| ID 20466 | KLRC1 | AT |
| ID 15868 | SERPING1 | AA |
| ID 69262 | DCUN1D4 | AP |
| ID 46814 | TNFAIP8L1 | AP |
| ID 54437 | KDM3A | ES |
| ID 3663 | ZNHIT6 | RI |
| ID 36590 | GPR56 | ES |
| ID 53528 | PPP1R21 | AA |
| ID 89646 | SH3KBP1 | ES |
| ID 36495 | MT1X | AD |
| ID 46854 | SAFB | ES |
| ID 46862 | RPL36 | RI |
| ID 42388 | ACSF2 | ES |
| ID 29072 | IFI27 | AP |
| ID 81426 | IMMP2L | AT |
| ID 51760 | VSTM1 | ME |
| ID 42788 | CLTC | ES |
| ID 76086 | OARD1 | ES |
| ID 88682 | SMS | ES |
| ID 43111 | PSMD12 | ES |
| ID 59128 | UQCC1 | ES |
Table 2
The most significantly related top 20 AS events.

| gene             | Pvalue     | HR          |
|------------------|------------|-------------|
| ZC3H11A-9457-AA  | 0.000101451| 4.56082E+19 |
| PLEKHG5-473-AA   | 0.000224123| 1.52E-13    |
| SCNN1B-35582-AA  | 0.000445583| 1.65E-17    |
| DLD-81379-AA     | 0.00070694 | 1.68E-09    |
| CDK20-86783-AA   | 0.000752571| 921.8005565 |
| RCHY1-69517-AA   | 0.001094207| 1.51E-16    |
| SUMO1-56937-AA   | 0.001336979| 1.08E-23    |
| RSRC2-24968-AA   | 0.00134264 | 15.22177842 |
| SETD5-63094-AA   | 0.001389135| 26.84300721 |
| PIF1-31133-AA    | 0.001432631| 2.33E-06    |
| MED16-46341-AA   | 0.001572585| 3.94134E+11 |
| PSM7-37561-AA    | 0.001691016| 0           |
| GATA2-6647-AA    | 0.001887582| 1.24E-15    |
| PHF20L1-85197-AA | 0.002055011| 0.005246581 |
| MAPK8-11475-AA   | 0.002144316| 590.8374194 |
| RNH1-13679-AA    | 0.002772059| 18.51886718 |
| USP19-64840-AA   | 0.002979425| 0.0207652   |
| PSMF1-58475-AA   | 0.003121573| 3405.819277 |
| IKBKG-90655-AA   | 0.00326978 | 0.014476702 |
| KIF12-87307-AA   | 0.003307827| 0.026646045 |
| RGPDB-54993-AD   | 0.00341962 | 7.69E-11    |
| HAUS4-26678-AD   | 0.00472909 | 77.68621278 |
| LCMT1-35669-AD   | 0.0083894  | Inf         |
| TFDP1-26387-AD   | 0.00884048 | 0.000214503 |
| SETD1B-14901-AD  | 0.00898551 | 1.86E-64    |
| SUV3H2-10832-AD  | 0.01124281 | 4.06E-05    |
| SNX19-19513-AD   | 0.01197651 | 4.05E-34    |
| PPP1CB-53075-AD  | 0.01485411 | 2.23E-11    |
| DCAF11-26842-AD  | 0.0152209 | 0.000940669 |
| PRMT5-26673-AD   | 0.01746529 | 411.6573381 |
| PLEKH1-28068-AD  | 0.01928892 | 0.009065525 |
| VCPKMD-27451-AD  | 0.02051333 | 0.106589638 |
| FAM86C1-17435-AD | 0.02411569 | 24.69611292 |
| GABARAP1-1-20354-AD| 0.03076203 | 0.237048957 |
| AMICA1-18973-AD  | 0.03076751 | 0.000105425 |
| PANK2-58614-AD   | 0.03382131 | 0.015836651 |
| MT1X-36495-AD    | 0.03841985 | 3.96477E+16 |
| CALM1-28825-AD   | 0.05148658 | 7.66E + 41  |
| HDHD2-45440-AD   | 0.0522931 | 34.6616706  |
| PCMT1-78109-AD   | 0.05401437 | 0.014100295 |
| CSTA-66429-AT    | 0.06204099 | 2.89E-12    |
| SSC5D-52113-AT   | 0.06327486 | 5.01E-22    |
| NXPE2-18845-AT   | 0.06474509 | 0.07672339 |
| CMPK2-52592-AT   | 0.065558  | 922.723720.5|
| ZNF233-50311-AT  | 0.06564533 | 0.004176995 |
| WHDM1-11493-AT   | 0.06755656 | 3.12E-14    |
| AMACR-71697-AT   | 0.0815728 | 6.63E-05    |
| PITPNM2-25042-AT | 0.01101959 | 0.002003646 |
| Gene Name | Value 1 | Value 2 |
|-----------|---------|---------|
| TRPV3-38443-AT | 0.001336752 | 0.223328548 |
| ANAPC7-24422-AT | 0.001485036 | 91.71026832 |
| ERICH1-82553-AT | 0.002255413 | 0.02666 |
| TUBCP3-26303-AT | 0.00245317 | 0.01430986 |
| NFX1-86100-AT | 0.002478114 | 38.63966959 |
| GREB1-52702-AT | 0.002938889 | 0.03820537 |
| WDR93-32441-AT | 0.003017199 | 4.684164517 |
| ABI3BP-65841-AT | 0.003554811 | 1.32E-05 |
| KLRK1-20466-AT | 0.003567846 | 5.90E-07 |
| IMP2L-81426-AT | 0.004225221 | 1.23E + 24 |
| ALS2-56917-AT | 0.004909742 | 100.7645844 |
| RNF150-70663-AT | 0.005071536 | 2.80E-06 |
| FAM175A-69797-AP | 0.000293772 | 1.65E-08 |
| ZFHX3-37557-AP | 0.000359674 | 0.085790031 |
| OSCP1-1775-AP | 0.000393564 | 1.85E-10 |
| FANK1-13439-AP | 0.000589627 | 0.000236832 |
| PPP1R14B-16606-AP | 0.000614074 | 1.33E-09 |
| CPD-40110-AP | 0.000625836 | 4.05E-10 |
| CPEB4-74588-AP | 0.000703225 | 8.63E-11 |
| NFKB2-12946-AP | 0.000773302 | 0.049398364 |
| TBC1D16-44009-AP | 0.000809099 | 4225.70548 |
| PRRL3-22026-AP | 0.000918145 | 3544.334838 |
| CYBRD1-55935-AP | 0.000939117 | 2.50E-09 |
| VAMP3-509-AP | 0.000940297 | 0.0203768 |
| FGYY-3207-AP | 0.00112074 | 0.000176264 |
| KHDRBS3-85269-AP | 0.001149967 | 1.59E-09 |
| MROH6-85423-AP | 0.001538554 | 4.629136 |
| STOML1-31619-AP | 0.001586055 | 0.009419939 |
| OGFOD2-25006-AP | 0.002632268 | 0.118023732 |
| SLC39A1-15734-AP | 0.002772756 | 4.06E-13 |
| BIN2-21837-AP | 0.003117258 | 0.155879643 |
| SLC3A1-53420-AP | 0.003560335 | 351.1571998 |
| VSTM1-51760-ME | 0.004226824 | 13.2699347 |
| SLC7A2-82807-ME | 0.035937025 | 389.8689632 |
| ANGPTL4-47225-RI | 0.001330367 | 1.07E-09 |
| EPHA10-1829-RI | 0.000374486 | 2.27E-13 |
| SLC45A4-85333-RI | 0.000521285 | 0.011000386 |
| ZNF83-51484-RI | 0.000564865 | 0.189810718 |
| UNC50-54642-RI | 0.000969924 | 0.022955033 |
| ADM-14339-RI | 0.001117022 | 2.54E-73 |
| NPIPB4-35513-RI | 0.001169873 | 0.057217512 |
| LCMT2-30228-RI | 0.001181364 | 1.47E-10 |
| PYCR1-44243-RI | 0.001187636 | 2.8881E + 16 |
| NPY1R-71011-RI | 0.001212912 | 0.00021434 |
| NPIPB5-35570-RI | 0.001229674 | 0.059093419 |
| COQ7-34285-RI | 0.001403404 | 3.46E-26 |
| KLRK1-85994-RI | 0.001439090 | 4.58E-31 |
| DNASE1-33608-RI | 0.001551965 | 0.144594873 |
| TPMRSS12-21715-RI | 0.001727746 | 0.151350654 |
| RPL30-84641-RI | 0.002101357 | 760966.3239 |
| DSP28-58266-RI | 0.002416703 | 11.79840842 |
| PARP10-85521-RI | 0.002516845 | 4483171.098 |
| HFC1-90543-RI | 0.002891349 | 4.10E-15 |
| WRAP53-39047-RI | 0.002977347 | 9.983272482 |
| APT7B-25959-ES | 0.000121302 | 3.85E-10 |
| TANGO2-61125-ES | 0.000144758 | 0.041051746 |
| RAD17-72360-ES | 0.000199573 | 0.000792792 |
| UPRT-B9523-ES | 0.000204142 | 115.2635627 |
| ZKSCAN1-80869-ES | 0.000215382 | 2.98E-08 |
| GK-88738-ES | 0.000296588 | 0.000184541 |
| SPIDR-83788-ES | 0.00031082 | 4.16E-05 |
| SNCA-69930-ES | 0.000346929 | 1.01E-10 |
| PPP2R4-87859-ES | 0.000368029 | 6.38E-09 |
| MAP4-64564-ES | 0.000406305 | 1.01E-10 |
| NEDD4L-45655-ES | 0.000527779 | 8.01E-65 |
| CORO7-33675-ES | 0.000549422 | 1.31E-08 |
| ITGB2-60850-ES | 0.000592232 | 1.05E-112 |
| PDLIM2-83027-ES | 0.000595418 | 5.40E + 34 |
| TAX1BP1-79065-ES | 0.000640984 | 5.74E + 78 |
| WWIP1-84363-ES | 0.000703433 | 2.63E-20 |
Molecular characteristics of survival-associated AS events
To reveal the molecular characteristics of these genes with survival associated-AS events, several bioinformatics analyses were conducted. According to the functional annotations of DAVID in Supplementary Table 2, “negative regulation of transcription from RNA polymerase II promoter”, “aging” and “cellular response to DNA damage stimulus” were the three most obvious biological process terms (Fig. 4A). “cytosol”, “nucleoplasm” and “extracellular exosome” were the three most significant cellular component terms (Fig. 4B). For molecular function, “protein binding”, “ligase activity” and “chromatin binding” were three most enriched categories (Fig. 4C). More importantly, we found that the “Ubiquitin mediated proteolysis” pathway was most significantly correlated with these genes (Fig. 4D).

Prognostic signatures for TNBC patients
We chose the most important survival-related AS events (excluding ME events, < 20) among the six types of AA, AD, AP, AT, ES, and RI as candidates in seeking for independent prognostic factors for TNBC patients. In order to eliminate the events that may not be independent factors in the prognosis gene models, the six types of candidate splicing events were applied by Multivariate Cox regression models respectively. The six signatures established with different types of AS events revealed a powerful ability to distinguish between improvement and deterioration in TNBC patients. Based on the six different types of AS in the TNBC cohort, the survival times of the two groups in each prognostic model were significantly different as shown in Fig. 5A-G. Furthermore, as presented in Fig. 5H, the AUC were significantly different in different splice type models. ROC curves were used to validate the efficiency of these prognostic models. Compared to other type of AS models, the ROC curve demonstrates that the ultimate use of all types of AS predictors has the highest efficiency in distinguishing between improvement and deterioration in TNBC patients (AUC 0.8). These independent prognostic and specific AS events were further in depth analyzed to establish the final prognostic predictors for TNBC in Table 3. The final prognostic signature was the most ideal predictor (Fig. 6A). This signature was easier and more valuable because it could well distinguish TNBC patients.
with distinct clinical outcomes in clinical practice (Fig. 6B).

### Table 3
Prognostic predictor for TNBC patients.

| ID   | Gene    | HR         | P value    | Type |
|------|---------|------------|------------|------|
| ID 7851 | PBXIP1  | -9.221E + 02 | 0.005759 ** | ES   |
| ID 60850 | ITGB2   | -7.834E + 02 | 1.59e-07 *** | ES   |
| ID 23661 | DCN     | -1.265E + 03 | 9.28e-05 *** | ES   |
| ID 29026 | UBR7    | -3.227E + 01 | 0.000111 *** | ES   |
| ID 64940 | RBM6    | -1.891E + 01 | 1.39e-05 *** | ES   |
| ID 31141 | ANKDD1A | -1.055E + 01 | 7.83e-06 *** | ES   |
| ID 42788 | CTLC    | -2.300E + 03 | 0.008895 ** | ES   |
| ID 43111 | PSMD12  | -1.091E + 03 | 6.36e-08 *** | ES   |
| ID 37561 | PSMD7   | -2.671E + 03 | 8.82e-07 *** | AA   |
| ID 11475 | MAPK8   | -1.092E + 01 | 0.001069 ** | AA   |
| ID 26387 | TFD51   | -8.987E + 00 | 0.033156 * | ES   |
| ID 19513 | SNX19   | -1.147E + 02 | 0.037601 * | AD   |
| ID 16606 | PPP1R14B | -2.198E + 01 | 0.026080 * | AP   |
| ID 22026 | PRR13   | 1.884E + 01  | 0.009380 ** | AP   |
| ID 55935 | CYBRD1  | -4.815E + 01 | 0.019633 * | AP   |
| ID 11493 | WDFY4   | -5.773E + 01 | 0.002559 ** | AT   |
| ID 82553 | ERICH1  | -1.211E + 01 | 4.63e-07 *** | AT   |
| ID 52702 | GREB1   | -5.658E + 00 | 0.009240 ** | AT   |
| ID 44243 | PYCR1   | 3.198E + 01  | 0.036101 * | RI   |
| ID 58266 | DUSP28  | 4.568E + 00  | 0.000471 *** | RI   |
| ID 42165 | NFE2L1  | -5.042E + 01 | 0.000940 *** | RI   |

### Construction of survival-associated AS events network

Nowadays, the phenomenon that the disordered AS events were mainly caused by many SFs has been widely recognized. Next, it was examined whether important portion of these AS events might be regulated by some key SFs that alter the expression of TNBC. A splicing regulatory network of 208 most significant survival-associated AS events in TNBC was characterized in Table 4. Using the genes expression levels calculated from the RNA sequencing data of level 3 of TNBC in TCGA, 11 SFs with significant genes expression levels in OS in TNBC were identified (P < 0.05) and possess the ability to predict the OS of TNBC patients (Fig. 7A, B). In the TNBC related network, the purple dots represented 11 survival-related SFs, and their expressions were significantly correlated with 12 survival-related AS events. The green dots indicated 7 AS events that were significantly related to the good patient survival (HR < 1), while the red dots indicated 5 AS events that were closely related to poor patient survival (HR > 1). Furthermore, the correlation between the genes expression levels of survival-related SFs and the PSI values of most survival-related AS events was calculated by the Spearman test. Related networks were built and only significant correlations as shown in Fig. 7C. We found an interesting phenomenon that most poor survival prognostic AS events were positively correlated (red lines) with the expression of SFs, while most good survival prognostic AS events were negatively correlated (green lines).
correlated (green lines) with the expression of SFs. The dot plots showed the correlation between splicing factor ALYREF and AD of MT1X, and the correlation of splicing factor MFSD11 and ES of ANKDD1A (Fig. 7D, E). The high expression of ALYREF was significantly correlated with low survival rate, while the high expression of MFSD11 was significantly related to the good survival of the patients.

Table 4
Survival associated SFs.

|   | Splicing factor |
|---|----------------|
| 1 | ZFR            |
| 2 | PPIE           |
| 3 | ALYREF         |
| 4 | RBFOX2         |
| 5 | HNRNPA3P1      |
| 6 | MFSD11         |
| 7 | WDR77          |
| 8 | BCAS2          |

Discussion

AS was a universal regulatory mechanism for gene expression that allowed a single gene to produce multiple unique mRNAs (Baralle & Giudice, 2017). A genome-wide study estimates that 90–95% of genes have experienced AS (Pan, Shai, Lee, Frey, & Blencowe, 2008). One of the molecular signs of cancer was AS abnormality (Ladomery, 2013). Previously, some studies exploring AS signatures have shown that aberrant AS guides many genes involved in cancer occurrence, transformation, and metastasis, and can be used as biomarkers and therapeutic targets for cancer diagnosis, prediction and prognosis (Oltean & Bates, 2014). In recent years, the development of bioinformatics methods and high-throughput sequencing technologies has helped to develop more reliable biomarker diagnostic methods (Gao, Zhong, Patel, Alur, & Vyas, 2017). These have provided effective help for fully understanding AS events in TNBC. Here we integrate the splicing map of TNBC patients with the clinical factors of TCGA to fully investigate the prognostic value of AS.

The evidence presented thus far supports that specific abnormal AS plays a significant part in the initiation, progression and metastasis of breast cancer. For example, ESRP1 and hnRNPM can alter the splicing pattern of exons in CD44 to produce different specific subtypes and correlate with the interstitial state of breast tumors. Among them, CD44v subtype is associated with TNBC carcinogenesis. It promotes the dryness of cancer cells by activating PDGFRβ/Stat3 cascade and
PFKFB4-mediated glucose metabolism, which is an important factor in tumor metastasis (Xu et al., 2014; H. Zhang et al., 2019). Similarly, the splicing factor SRSF1, which is up-regulated in human breast tumor cells, promotes AS of the tumor gene MDM2 to produce a MDM2-ALT1 subtype with tumorigenic properties (Comiskey, Jacob, Singh, Tapia-Santos, & Chandler, 2015). Due to the importance of AS in cancer biology, more and more studies have focused on the clinical relevance of AS in malignant tumors.

In this study, we used several biomedical analysis methods to integrate AS event profiles and clinical information from TNBC patients into prognostic-associated AS. A splicing prognostic marker that can divide TNBC patients into subgroups with different survival outcomes was constructed. First of all, the AS signatures of 115 patients with TNBC were analyzed, followed by a comprehensive survival analysis and a powerful prognostic predictor. Nearly half of the AS events (1,428 AS events from 1,383 genes) were ES, indicating that AS is common phenomenon in TNBC, and ES is the most common type of splicing. This shows that AS has great potential in applications. ANAPC7 (APC7) and Nedd4L are central genes in PPI network analysis. It is worth noting that APC has been identified as essential for the pathogenesis of breast cancer (Khan, Arafah, Shaik, Mahale, & Alanazi, 2018). Some reports show APC plays a central part in inhibiting Wnt signaling pathways that control TNBC cell proliferation and differentiation (De et al., 2016; Lang et al., 2017; Lv et al., 2019). Guarnieri AL presented novel findings indicating that Nedd4L had a very important role in suppressing breast cancer (Guarnieri et al., 2018). In addition, we sought to study the underlying mechanisms of prognostic AS events in TNBC. It is worth noting that the transcription of these genes can be modified through exosomes-related pathways to modify protein markers, thereby participating in processes such as cell proliferation, migration and apoptosis. This was discovered from the CC aspect of the GO analysis in our current work. Functional enrichment analysis indicated that the ubiquitin-mediated proteolytic pathway is an important pathway of interference, consistent with studies of AS in breast and colorectal cancer (Xiong et al., 2018; Zhang, Duan, Cun, & Yang, 2019).

Based on the overall survival of the AS events, the prognostic features are the focus of our current study to facilitate monitoring the prognosis of patients with TNBC. Studies have shown that a variety
of molecules can be used as a special diagnosis and independent prognostic marker for tumors. Recently, many biomarkers for breast cancer have been developed. A study developed a prognostic signature for the 19 gene associated with clinical prognosis in patients with BRCA (Su, Miao, Ye, Cui, & He, 2019). In addition, BCL2 and CD82 are both considered as potential and reliable biomarkers for breast cancer diagnosis (Wang et al., 2019). For TNBC, which is more difficult to cure, studies have also shown that there are related prognostic markers (Cai et al., 2019).

Here, we used a multivariate Cox regression model to screen out a series of AS events to promote clinical metastasis. The prognostic markers we present show an ideal performance in predicting the patient's clinical outcome. Ultimately, we achieved a combination of all available types of AS. The AUC of ROC for final prognostic predictor was 0.8, which was much higher than all prognostic indicators established using only one type of AS, indicating that the application of predictive power enhances the prognostic potential of TNBC patients. SF is an important regulator of AS events. Changes in AS events occur in a variety of tumors, suggesting that SFs may be important molecules in splicing disorders in cancer (Climente-Gonzalez, Porta-Pardo, Godzik, & Eyras, 2017). More and more people believe that many SFs have changed and participated in the development of BRCA cancer cells through various mechanisms (Anczukow et al., 2012; Dolfini, Andrioletti, & Mantovani, 2019; Gokmen-Polar et al., 2019). In our study, we not only explored the relationship between AS events and tumorigenesis, but also emphasized the potential role of SF in TNBC. A potential SF-AS related network was constructed between prognostic SF and the most significant survival-related AS events. AS events with good prognosis were positively correlated with splicing factor expression, while AS events with poor prognosis were negatively correlated with splicing factor expression. For example, ALYREF is an RNA binding protein that ligates to transcription. It has been reported that ALYREF may not only be a molecular marker for early detection of lymph node metastasis, but may also be effective in preventing oral squamous cell carcinoma metastasis (Saito et al., 2013). However, the role of ALYREF in TNBC has not been explored. Therefore, whether the down-regulation of some specific SFs leads to a reduction in favorable prognostic AS events and an increase in poor prognosis AS events requires further verification by functional experiments. But we have raised important
questions about the potential key SFs in TNBC. Up-regulation or down-regulation of SFs expression may lead to abnormal splicing and differential expression of splice variants. Upregulation of certain oncogenic SFs could promote TNBC progression. Because of the exploration of prognostic prediction models, more personalized methods for different patients could precisely target and regulate AS in TNBC patients, providing a wealth of biomarker candidates and potential targets for TNBC treatment.

Conclusions
In summary, we revealed the system characteristics of AS events in TNBC, and found that our final model of prognostic indicators related to survivor-related AS events performed well in the risk stratification of TNBC patients, which is promising in clinical applications. In addition, we also found a unique splicing-related network in TNBC patients. These results revealed the role of RNA splicing in the development of TNBC, which is the most valuable and meaningful. In-depth analysis of RNA splicing patterns may indeed reveal new drivers of cancer, providing valuable therapeutic targets for future validation of TNBC.

Appendices
Additional file:
Supplementary Table 1. Collection of 404 SFs.
Supplementary Table 2. Detailed enrichment results of GO terms and KEGG pathways.
Supplementary Table 3. Correlation between SFs expression profile and OS.
Supplementary Table 4. Correlation between SFs expression and PSI of AS events.

Abbreviations
TNBC: triple-negative breast cancer; AS: Alternative splicing; TCGA: The Cancer Genome Atlas; SF: Splicing factor; BC: Breast cancer; ER: Estrogen receptor; PR: Progesterone receptor; HER2: Human epidermal growth factor receptor 2; PSI: Percent Spliced In; ES: Exon Skip; ME: Mutually Exclusive Exons; RI: Retained Intron (RI); AP: Alternate Promoter; AT: Alternate Terminator; AD: Alternate Donor site; AA: Alternate Acceptor site; ROC: Receiver Operating Characteristic Curve; AUC: Area Under Curve; OS: Overall survival; PPI: protein-protein interaction; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; MF: molecular functions; BP: biological processes; CC: cellular components.
Declarations

Acknowledgements

Not applicable.

Web links and URLs

The Cancer Genome Atlas database (https://portal.gdc.cancer.gov/repository).

Database for Annotation, Visualization and Integration Discovery (DAVID) (Discovery https://david.ncifcrf.gov/).

Cytoscape (https://cytoscape.org/).

Authors’ contributions

SQZ, SG, MJW and MH conceived and designed the study. XML, RM, YFZ, YYY, JW, XYS, MCG, DD, HLZ and LH, collected and processed data. SQZ and SG analyzed data. SQZ prepared tables and figures.

SQZ drafted the manuscript. MJW and MH revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by National Natural Science Foundation of China and Liaoning joint fund key program (No.U1608281), National Natural Science Foundation of China (NSFC, No. 81972794, 81902708), Key R&D Guidance Plan Projects in Liaoning Province (2019JH8/10300011), Shenyang S&T Projects (19-109-4-09), Liaoning provincial department of education scifentific research project (QN2019034) and China Medical University Start up fund for postdoctoral research No.1210618017).

Availability of data and materials

The datasets generated and/or analysed during the current study are available in The Cancer Genome Atlas database and additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details

1Department of Pharmacology, School of Pharmacy, China Medical University, Shenyang, Liaoning Province, China

2Liaoning Key Laboratory of molecular targeted anti-tumor drug development and evaluation; Liaoning Cancer immune peptide drug Engineering Technology Research Center; Key Laboratory of Precision Diagnosis and Treatment of Gastrointestinal Tumors, Ministry of Education; Shenyang, Liaoning Province, China

3Department of Breast Surgery, First Affiliated Hospital of China Medical University, Shenyang, China.

References

1. Anczukow, O., Rosenberg, A. Z., Akerman, M., Das, S., Zhan, L., Karni, R., . . . Krainer, A. R. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat Struct Mol Biol, 19*(2), 220-228. doi:10.1038/nsmb.2207

2. Baralle, F. E., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol, 18*(7), 437-451. doi:10.1038/nrm.2017.27

3. Cai, P., Lu, Z., Wu, J., Qin, X., Wang, Z., Zhang, Z., . . . Zhao, J. (2019). BTN3A2 serves as a prognostic marker and favors immune infiltration in triple-negative breast cancer. *J Cell Biochem*. doi:10.1002/jcb.29485

4. Climente-Gonzalez, H., Porta-Pardo, E., Godzik, A., & Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. *Cell Rep, 20*(9), 2215-2226. doi:10.1016/j.celrep.2017.08.012

5. Comiskey, D. F., Jr., Jacob, A. G., Singh, R. K., Tapia-Santos, A. S., & Chandler, D. S. (2015). Splicing factor SRSF1 negatively regulates alternative splicing of MDM2 under damage. *Nucleic Acids Res, 43*(8), 4202-4218. doi:10.1093/nar/gkv223
6. Criscitiello, C., Bayar, M. A., Curigliano, G., Symmans, F. W., Desmedt, C., Bonnefoi, H., . . . Andre, F. (2018). A gene signature to predict high tumor-infiltrating lymphocytes after neoadjuvant chemotherapy and outcome in patients with triple-negative breast cancer. *Ann Oncol*, 29(1), 162-169. doi:10.1093/annonc/mdx691

7. David, C. J., & Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev*, 24(21), 2343-2364. doi:10.1101/gad.1973010

8. de Nonneville, A., Finetti, P., Adelaide, J., Lambaudie, E., Viens, P., Goncalves, A., . . . Bertucci, F. (2019). A Tyrosine Kinase Expression Signature Predicts the Post-Operative Clinical Outcome in Triple Negative Breast Cancers. *Cancers (Basel)*, 11(8). doi:10.3390/cancers11081158

9. De, P., Carlson, J. H., Wu, H., Marcus, A., Leyland-Jones, B., & Dey, N. (2016). Wnt-beta-catenin pathway signals metastasis-associated tumor cell phenotypes in triple negative breast cancers. *Oncotarget*, 7(28), 43124-43149. doi:10.18632/oncotarget.8988

10. Dolfini, D., Andrioletti, V., & Mantovani, R. (2019). Overexpression and alternative splicing of NF-YA in breast cancer. *Sci Rep*, 9(1), 12955. doi:10.1038/s41598-019-49297-5

11. Gao, M., Zhong, A., Patel, N., Alur, C., & Vyas, D. (2017). High throughput RNA sequencing utility for diagnosis and prognosis in colon diseases. *World J Gastroenterol*, 23(16), 2819-2825. doi:10.3748/wjg.v23.i16.2819

12. Genov, N., Basti, A., Abreu, M., Astaburuaga, R., & Relogio, A. (2019). A bioinformatic analysis identifies circadian expression of splicing factors and time-dependent alternative splicing events in the HD-MY-Z cell line. *Sci Rep*, 9(1), 11062. doi:10.1038/s41598-019-47343-w
13. Gokmen-Polar, Y., Neelamraju, Y., Goswami, C. P., Gu, Y., Gu, X., Nallamothu, G., ... Badve, S. S. (2019). Splicing factor ESRP1 controls ER-positive breast cancer by altering metabolic pathways. EMBO Rep, 20(2). doi:10.15252/embr.201846078

14. Guarnieri, A. L., Towers, C. G., Drasin, D. J., Oliphant, M. U. J., Andrysik, Z., Hotz, T. J., ... Ford, H. L. (2018). The miR-106b-25 cluster mediates breast tumor initiation through activation of NOTCH1 via direct repression of NEDD4L. Oncogene, 37(28), 3879-3893. doi:10.1038/s41388-018-0239-7

15. Khan, Z., Arafah, M., Shaik, J. P., Mahale, A., & Alanazi, M. S. (2018). High-frequency deregulated expression of Wnt signaling pathway members in breast carcinomas. Onco Targets Ther, 11, 323-335. doi:10.2147/OTT.S154395

16. Ladomery, M. (2013). Aberrant alternative splicing is another hallmark of cancer. Int J Cell Biol, 2013, 463786. doi:10.1155/2013/463786

17. Lang, M., Berry, D., Passecker, K., Mesteri, I., Bhuju, S., Ebner, F., ... Gasche, C. (2017). HuR Small-Molecule Inhibitor Elicits Differential Effects in Adenomatosis Polyposis and Colorectal Carcinogenesis. Cancer Res, 77(9), 2424-2438. doi:10.1158/0008-5472.CAN-15-1726

18. Lee, S. C., & Abdel-Wahab, O. (2016). Therapeutic targeting of splicing in cancer. Nat Med, 22(9), 976-986. doi:10.1038/nm.4165

19. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph, 20(12), 1983-1992. doi:10.1109/TVCG.2014.2346248

20. Li, Y., Sun, N., Lu, Z., Sun, S., Huang, J., Chen, Z., & He, J. (2017). Prognostic alternative mRNA splicing signature in non-small cell lung cancer. Cancer Lett, 393, 40-51. doi:10.1016/j.canlet.2017.02.016

21. Lin, P., He, R. Q., Huang, Z. G., Zhang, R., Wu, H. Y., Shi, L., ... He, Y. (2019). Role
of global aberrant alternative splicing events in papillary thyroid cancer prognosis. 

*Aging (Albany NY)*, 11(7), 2082-2097. doi:10.18632/aging.101902

22. Lu, Z. X., Huang, Q., Park, J. W., Shen, S., Lin, L., Tokheim, C. J., . . . Xing, Y. (2015). Transcriptome-wide landscape of pre-mRNA alternative splicing associated with metastatic colonization. *Mol Cancer Res*, 13(2), 305-318. doi:10.1158/1541-7786.MCR-14-0366

23. Lv, Z. D., Xin, H. N., Yang, Z. C., Wang, W. J., Dong, J. J., Jin, L. Y., & Li, F. N. (2019). miR-135b promotes proliferation and metastasis by targeting APC in triple-negative breast cancer. *J Cell Physiol*, 234(7), 10819-10826. doi:10.1002/jcp.27906

24. Oltean, S., & Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene*, 33(46), 5311-5318. doi:10.1038/onc.2013.533

25. Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 40(12), 1413-1415. doi:10.1038/ng.259

26. Ryan, M. C., Cleland, J., Kim, R., Wong, W. C., & Weinstein, J. N. (2012). SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics*, 28(18), 2385-2387. doi:10.1093/bioinformatics/bts452

27. Saito, Y., Kasamatsu, A., Yamamoto, A., Shimizu, T., Yokoe, H., Sakamoto, Y., . . . Uzawa, K. (2013). ALY as a potential contributor to metastasis in human oral squamous cell carcinoma. *J Cancer Res Clin Oncol*, 139(4), 585-594. doi:10.1007/s00432-012-1361-5

28. Seiler, M., Peng, S., Agrawal, A. A., Palacino, J., Teng, T., Zhu, P., . . . Yu, L. (2018). Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep*, 23(1), 282-296 e284.
29. Sherman, B. T., Huang da, W., Tan, Q., Guo, Y., Bour, S., Liu, D., . . . Lempicki, R. A. (2007). DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics, 8*, 426. doi:10.1186/1471-2105-8-426

30. Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J Clin, 70*(1), 7-30. doi:10.3322/caac.21590

31. Su, J., Miao, L. F., Ye, X. H., Cui, M. S., & He, X. F. (2019). Development of prognostic signature and nomogram for patients with breast cancer. *Medicine (Baltimore), 98*(11), e14617. doi:10.1097/MD.00000000000014617

32. Trincado, J. L., Sebestyen, E., Pages, A., & Eyras, E. (2016). The prognostic potential of alternative transcript isoforms across human tumors. *Genome Med, 8*(1), 85. doi:10.1186/s13073-016-0339-3

33. Wang, X., Zhong, W., Bu, J., Li, Y., Li, R., Nie, R., . . . Li, Y. (2019). Exosomal protein CD82 as a diagnostic biomarker for precision medicine for breast cancer. *Mol Carcinog, 58*(5), 674-685. doi:10.1002/mc.22960

34. Xiong, Y., Deng, Y., Wang, K., Zhou, H., Zheng, X., Si, L., & Fu, Z. (2018). Profiles of alternative splicing in colorectal cancer and their clinical significance: A study based on large-scale sequencing data. *EBioMedicine, 36*, 183-195. doi:10.1016/j.ebiom.2018.09.021

35. Xu, Y., Gao, X. D., Lee, J. H., Huang, H., Tan, H., Ahn, J., . . . Cheng, C. (2014). Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev, 28*(11), 1191-1203. doi:10.1101/gad.241968.114

36. Zhang, D., Duan, Y., Cun, J., & Yang, Q. (2019). Identification of Prognostic Alternative Splicing Signature in Breast Carcinoma. *Front Genet, 10*, 278.
doi:10.3389/fgene.2019.00278

37. Zhang, D., Duan, Y., Wang, Z., & Lin, J. (2019). Systematic profiling of a novel prognostic alternative splicing signature in hepatocellular carcinoma. *Oncol Rep*, 42(6), 2450-2472. doi:10.3892/or.2019.7342

38. Zhang, H., Brown, R. L., Wei, Y., Zhao, P., Liu, S., Liu, X., . . . Cheng, C. (2019). CD44 splice isoform switching determines breast cancer stem cell state. *Genes Dev*, 33(3-4), 166-179. doi:10.1101/gad.319889.118

39. Zhu, J., Chen, Z., & Yong, L. (2018). Systematic profiling of alternative splicing signature reveals prognostic predictor for ovarian cancer. *Gynecol Oncol*, 148(2), 368-374. doi:10.1016/j.ygyno.2017.11.028

40. Zong, Z., Li, H., Yi, C., Ying, H., Zhu, Z., & Wang, H. (2018). Genome-Wide Profiling of Prognostic Alternative Splicing Signature in Colorectal Cancer. *Front Oncol*, 8, 537. doi:10.3389/fonc.2018.00537

Tables
Due to technical limitations, Tables 1-4 are provided in the Supplementary Files section.

Figures
Figure 1

Flowchart for profiling AS of TNBC.
Figure 2

Overview of AS events profiling in TNBC cohort. A, Illustration for splicing pattern of seven sorts of AS events, including Exon Skip (ES), Alternate Promoter (AP), Alternate Terminator (AT), Alternate Donor site (AD), Alternate Acceptor site (AA), Mutually Exclusive Exons (ME),
and Retained Intron (RI). B, The number of AS events and corresponding genes included in the present study. C, The number of prognosis-related AS events and corresponding genes obtained by using univariate COX analysis. D, UpSet plots in TNBC, showing the interactions among the seven types of AS events. One gene may have up to three types of AS events.

Figure 3

Gene network of survival associated AS in TNBC generated by Cytoscape. The bigger the point in the network, the more important it is.
Functional analyses on parent genes from survival-related AS events in TNBC, including GO and KEGG. A, biological processes. B, Cellular component. C, molecular function. The dot size represents the enriched gene number. D, KEGG pathway analysis of genes with survival-associated AS events.
Kaplan-Meier plots and ROC curves of prognostic predictors for TNBC patients. A to F, Kaplan-Meier curves of prognostic predictors built with one type of AS events for TNBC patients, respectively. G, Kaplan-Meier curves of the final prognostic predictors built with all types of AS events for TNBC patients. Red line indicates high risk group while green line indicates low risk group. H, ROC curves with AUCs of prognostic predictors built by one type or all seven types of AS events in TNBC.
Identification capability of prognostic signature for separating patients based on patients’ survival status and survival times into high-risk and low-risk groups.
Figure 7

Survival associated SFs and splicing correlation network in TNBC. A, Survival curves of prognostic splicing factor MFSD11 in TNBC. B, Survival curves of prognostic splicing factor ALYREF in TNBC. C, Splicing correlation network in TNBC. AS events whose PSI values were positively/negatively correlated with survival times were represented with green/red dots. Purple dots were survival associated SFs. The positive/negative correlation between expressions of SFs and PSI values of AS were represented with red/green lines. D, Dot plot of correlation between expression of MFSD11 and ES PSI values of ANKDD1A. E, Dot plot of correlation between expression of MTIX and AD PSI values of ALYREF.

Supplementary Files

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

Table 4.xlsx
Table 3.xlsx
Supplementary Table 4.xlsx
Supplementary Table 3.xlsx
Table 1.xlsx
Table 2.xlsx
Supplementary Table 1.xlsx
Supplementary Table 2.xlsx