

Analysis of relapse-associated alternative mRNA splicing and construction of a prognostic signature predicting relapse in I–III colon cancer

Zhiyuan Zhang

Zhongshan Hospital Fudan University <https://orcid.org/0000-0003-0336-7783>

Qingyang Feng

Zhongshan Hospital Fudan University

Peng Zheng

Zhongshan Hospital Fudan University

Yang Lv

Zhongshan Hospital Fudan University

Yihao Mao

Zhongshan Hospital Fudan University

Yuqiu Xu

Zhongshan Hospital Fudan University

Guodong He

Zhongshan Hospital Fudan University

Jianmin Xu (✉ zsyyxjm@163.com)

<https://orcid.org/0000-0002-5462-8421>

Primary research

Keywords: colon cancer, alternative-splicing events, relapse

Posted Date: March 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-17501/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Genomics on November 1st, 2020. See the published version at <https://doi.org/10.1016/j.ygeno.2020.07.002>.

Abstract

Background : The literature depicting the effects of alternative splicing (AS) events on relapse of colon cancer is little and there is no signature based on the alternative splicing.

Methods : The bioinformatic analysis was performed based on data of The Cancer Genome Atlas (TCGA) to identify the relapse-associated ASs, the potential interactions were further analyzed and a robust signature was built after univariate Cox regression, LASSO Cox regression, and multivariate Cox regression analysis to predict the relapse in I–III colon cancer. Molecular subtypes was identified based on the signature.

Results : We identified 1912 ASs of 1384 mRNA, based on the relapse-associated ASs, we constructed the network of protein-protein interactions (PPI) and ASs-splicing factors (SF) interactions. 1294 of proteins with 7396 interactions were included in the PPI network. 14 SFs combined with 78 relapse-associated ASs were included in the AS-SF network. We finally built a robust signature to predict the relapse of I–III colon cancer with a considerable AUC value in both the training group and the test group (0.857,0.839). Based on the ASs involved in the signature, samples were classified into 4 molecular subgroups distinguishing the relapse rate in diverse groups.

Conclusion : Our study provides a profile of relapse-associated ASs in I–III colon cancer and build a robust signature to predict the relapse of I–III colon cancer patients and further classify the patients into 4 molecular subtypes.

Background

Colorectal cancer (CRC) is one common type of cancer with the estimated third-place mortality and morbidity among all cancers in the prediction data of cancers in 2020[1]. Colon cancer which is different in terms of anatomy, treatment and molecular characterization from rectal cancer forms an important part of CRC. Surgical resection is the only curative treatment for colon cancer, and if combined with different chemotherapy strategies, the prognosis of colon cancer patients can be moderately improved[2]. Although the treatment and management of colon cancer have progressed greatly, patients still face many challenges. The relapse after surgery is a major concern that can induce poor prognosis among colon cancer patients. Previous data indicated that approximately 25–40% of patients could suffer recurrence or metastases after the surgical resection within 2 years, of which two-third was metastasis and one-third was local relapse[3–5]. Relevant data revealed that 90% relapse occurred within the first 3 years after surgery[6]. Currently, the most commonly used risk factors to predict the relapse are still tumor-node-metastasis (TNM) system, however, the TNM system cannot adequately distinguish the patients with high or low risk of recurrence. Treatment after relapse usually get a poor response due to the poor pathological features. Thus, the discovery of efficient predictive factors for relapse after surgery is urgent to improve the long-term survival of colon cancer patients.

Alternative splicing (AS) is a common mechanism to regulate the translation of diverse mRNA isoforms and to generate the different isoforms of protein from a single gene[7]. Almost 95% of genes undergo AS and AS plays a significant role in the process of regulating different biological functions in eukaryotic[8]. However, aberrant alternative splicing also can cause the occurrence and progression of cancers[9, 10]. AS can influence the expression of oncogenes or tumor suppressors and further result in the progression or the relapse of tumors[11]. In the field of CRC, many types of researches also proved the functions of AS: SRSF6 could regulate the AS to promote the tumor progression and HNRNPLL was recognized as a CRC suppressor[12, 13]. However, few types of researches reported the profiles of aberrant ASs in the field of the relapse of colon cancer and no one builds the signature with the aberrant AS to predict the relapse of colon cancer.

In this present study, we analyzed the aberrant ASs that were associated with relapse in I–III colon cancer and further built a robust signature to predict the relapse in I–III colon cancer patients based on the data The Cancer Genome Atlas (TCGA).

Materials And Methods

Data acquisition

RNA sequence data and relevant clinical information of colon cancer were obtained from The Cancer Genome Atlas (TCGA, <https://tcga-data.nci.nih.gov/tcga/>). TCGA is a landmark cancer genomics program containing over 20000 primary cancer and matched normal samples spanning 33 cancer types. The cohort contained RNA sequence data of 521 samples and relevant clinical information including the gender, age, tumor stage, preoperative treatment and information of relapse. Among the 521 samples, 393 colon cancer patients were identified as I–III stage. The alternative splicing events of colon cancer were downloaded from the TCGA SpliceSeq (<https://bioinformatics.mdanderson.org/TCGASpliceSeq/>). TCGA SpliceSeq is a resource to research the alterations in mRNA splicing patterns based on TCGA RNASeq data. The Percent Spliced In (PSI) value, which is recognized as the normalized way and ranges from 0 to 1 was utilized to evaluate the ASs. The filter was set as the percentage of samples with PSI value ≥ 75 to obtain reliable information of ASs. AS was recognized as 7 types including Alternate Acceptor site (AA), Alternate Donor site (AD), Alternate Promoter (AP), Alternate Terminator (AT), Exon Skip (ES), Mutually Exclusive Exons (ME), and Retained Intron (RI) (Fig. 1A). The list of relevant splicing factors (SFs) was downloaded from SpliceAcid-F, a database containing validated SFs[14]. The data processing was performed by R language (Version 3.6.2) and Perl (Version 5.30.1).

Identification of relapse-associated ASs in I–III colon cancer

In order to obtain the relapse-associated ASs in I–III colon cancer, we divided the colon cancer patients into 2 groups, relapse group, and long-term survival group. The relapse group was defined as patients who suffer recurrence or metastasis after surgery. Long-term survival group was defined as patients without relapse nor metastasis after a minimum of 3 years follow-up. Propensity score matching (PSM) analysis was performed between the 2 groups, tumor stage and adjuvant chemotherapy which were the

most influential factors associated with relapse were selected as matching options. The 2 groups were matched 1:1. The relapse-associated ASs were identified by Wilcoxon tests by comparing the 2 groups, P-value < 0.05 was considered as statistically significant. Results of relapse-associated ASs were shown as an UpSet plot.

Construction of protein-protein interaction network and SFs-ASs correlation network

The proteins which were involved in relapse-associated ASs were further analyzed by constructing a protein-protein interaction (PPI) network based on the online database STRING (<https://string-db.org/>, Version 11.0) database. The PPI network was analyzed by MCODE (for clustering the PPI network) and Cytohubba (for selection of hub proteins) in Cytoscape (Version 3.5.0). The correlation between PSI value of relapse-associated ASs and expression of SFs was analyzed by Spearman tests through R language (Version 3.6.2), P-value < 0.01 and Correlation Coefficient (cor) > 0.65 or <-0.65 was considered as statistically significant. Data on the PPI network and SFs-ASs correlation network were collected and further visualized through Cytoscape (Version 3.5.0).

Construction of the relapse prediction signature based on relapse-associated ASs.

The I–III colon cancer cohort from TCGA was divided into 2 groups randomly: the training group and the test group with the ratio 7:3. The training group was utilized to construct the relapse-associated ASs signature to predict the relapse in colon cancers and the test group was utilized to validate the accuracy. The relationship between relapse-associated ASs and relapse-free survival was analyzed by univariate Cox regression model. P-value < 0.05 was considered as statistically significant. The results of these ASs were visualized in terms of volcano plot, UpSet plot, and bubble plot through R language (Version 3.6.2). In order to make the signature more robust, these relapse-associated ASs with P-value < 0.001 in the univariate Cox regression model were selected into the next analysis. LASSO Cox regression model further filtered the relevant ASs and then Multivariate cox regression analysis was performed, the ASs with P < 0.05 under Multivariate cox regression analysis was carried out to build the signature. Finally, the accuracy of the signature was validated in the test group. The accuracy of the signature was tested through the ROC curve, risk score analysis, and relapse-free survival analysis. Univariate, Multivariate Cox regression and LASSO Cox regression analysis were performed by R language (Version 3.6.2).

Estimation of the relationship between molecular subtypes and clinical features.

Based on the PSI value of AS involved in the signature, the samples were classified. The classification was conducted by R package ConsensusClusterPlus. The optional number of clusters was chosen according to the Elbow method and the Gap statistic. The consensus molecular subtype was obtained through the CMScaller package. The association between clinical features was shown as the heatmap and Kaplan-Meier plot

Results

Identification of relapse-associated ASs

Data of PSI value from 484 colon cancer samples and data of RNA sequence from 521 samples with relevant clinical data were obtained from the online database, among these samples, 393 patients were identified as I–III stage. 68 out of the 393 patients were identified as the relapse group and 76 samples were identified as long-term survival group. PSM analysis was performed based on the clinical data of patients from the 2 groups. These 2 groups were matched by tumor stage and adjuvant chemotherapy. The clinical characteristics before and after PSM were shown in Table 1. Before PSM, distribution of the tumor stages in the 2 groups was statistically significant ($P = 0.04$), and after PSM, distribution of the tumor stage, as well as other clinical characteristics (age, gender, adjuvant chemotherapy) were all balanced. PSI value of the relevant patients in the 2 groups was compared to identify the aberrantly expressed ASs between the 2 groups, and these ASs were considered as relapse-associated ASs. Finally, 1912 ASs of 1384 mRNA were identified. The profile of the relapse-associated ASs was presented in terms of an UpSet plot to depict the overlapping of 7 subtypes ASs and related mRNA (Fig. 1B). The most frequent ASs type was ES (506) and the least frequent ASs was ME (10).

Table 1

Clinical characteristics of patients in relapse and long-term survival groups before and after PSM.

Variable	Before matching		P	After matching		P
	Relapse	Long-term survival		Relapse	Long-term survival	
Age(mean,range)	66.43(34–90)	68.8(37–90)	0.27	66.43(34–90)	68.(37–90)	0.33
Gender			0.37			0.49
Male	40	39		40	35	
Female	28	37		28	33	
Stage			0.04*			0.12
I	6	11		6	6	
II	30	46		30	43	
III	29	18		29	18	
NA	3	1		3	1	
Adjuvant chemotherapy			0.13			0.26
No	45	60		45	52	
Yes	23	16		23	16	
Total	68	76		68	68	
* indicates P-value < 0.05,the result is statistically significant.						

Construction of PPI network and SFs-ASs correlation network

Relapse-associated ASs can produce the diverse isoforms of proteins which may further result in the relapse of colon cancer. Thus, proteins that were involved in the relapse-associated ASs were selected for PPI network construction for further analysis. A number of 1294 of proteins with 7396 interactions were included in the PPI network, MCODE was utilized to classify the clusters and 30 clusters were identified, the scores for individual clusters were from 2.6–28.2. The cluster with the highest scores (28.2) was presented in Fig. 2A. In this cluster, 57 proteins and 789 interactions are included. Among these proteins, 10 proteins are identified as hub proteins by Cytohubba in this cluster (RBBP6, UBE3D, UBE2D3, HERC4, ARIH2, TRIP12, HERC6, UBE2D4, RCHY1, and SRSF3). RBBP6 ranked the highest among these 10 hub proteins. UBE2D3 tied for the second with RCHY1, the others all ranked the fourth place (Fig. 2B).

The correlation between relapse-associated ASs and SFs were also analyzed. 71 SFs which were validated by experimental researches were included in analysis. The interaction between relapse-associated ASs and SFs was analyzed by Spearman tests. Finally, 14 SFs, 78 relapse-associated ASs with a total of 102 interactions were involved in the network. DDX39B, PNN, ZC3H11A, MATR3, and PPIG were considered to be the core SFs that owned the most interactions with ASs. ETFB-51319-AP, ETFB-51320-AP, EIF2B1-25097-AT, EIF2B1-25098-AT, and RER1-256-AP were identified as the core relapse-associated ASs in the network (Fig. 3).

Univariate Cox regression and LASSO regression analysis

Samples with corresponding clinical data were divided into training and test group randomly, the training group was utilized to construct the signature. In the training group, the relapse-associated ASs were then further analyzed by univariate Cox regression. 986 ASs of 726 mRNA were considered as statistically significant ($P < 0.05$). The volcano plot (Fig. 4A) depicted the ratio of ASs with the absolute value of Z-score > 2 and $P < 0.05$. The overlapping profile of ASs and mRNA was shown in Fig. 4B. The most frequent ASs was AT and the least common ASs was ME. Bubble plot (Fig. 5) indicated the representative ASs with high Z-score and strict P-value. In RI, CGGBP1-65667-RI, ALS2CL-64463-RI, and EED-18179-RI ranked the highest. In ME, CHEK2-61534-ME, GCDH-47884-ME, and TMEM104-217418-ME occupied the top three. In AP, ARAP1-17637-AP, ANKRD54-62163-AP, and ANKRD54-62161-AP were the top three. In AT, BTNL8-75028-AT, AIG1-77972-AT, and ALG10-21064-AT ranked the highest. In ES, the top three were BCLAF1-77908-ES, ARFGAP1-60111-ES, and ARIH2-64787-ES. In AD, the top three were HDGF-8323-AD, CEP68-53781-AD, and CCDC84-19052-AD. In AA, BCL2L14-20499-AA, CGGBP1-65668-AA, and IL17RE-63243-AA took the first three places. The ASs with strict statistically significant value in univariate Cox regression were included in the LASSO Cox regression model. Cross-validation was conducted (Fig. 6A), the $\log(l)$ with the lowest deviance was picked to selected the ASs whose coefficients were not 0 in this regression model (Fig. 6B). Finally, 23 ASs were filtered through the LASSO Cox regression model.

Construction of relapse prediction signature of ASs and validation

Multivariate Cox regression analysis was performed based on the results of LASSO Cox regression analysis. Finally, 10 ASs were included in the signature, risk score= $(-15.25 \times \text{PSI value of CCNYL1-57200-AT}) + (5.26 \times \text{PSI value of CREM-11231-AP}) + (2.19 \times \text{PSI value of DECR1-84409-ES}) + (3.47 \times \text{PSI value of HDGF-8323-AD}) + (-4.20 \times \text{PSI value of POM121-79925-AP}) + (-26.63 \times \text{PSI value of SIRT3-13599-ES}) + (4.99 \times \text{PSI value of SLC15A4-25184-AP}) + (-4.19 \times \text{PSI value of TEX264-65099-AD}) + (5.24 \times \text{PSI value of ZBED5-14401-AA}) + (2.71 \times \text{PSI value of ZBTB7A-46761-AP})$. The detailed information of these ASs was shown in Table 2. Based on the signature, the risk score of each patient was calculated, and patients were divided into high-risk score group and low-risk score group according to the median risk score. Overall-survival and risk score analysis indicated that the signature could distinguish the 2 groups efficiently in the training group (Fig. 7A,B), and the ROC curve of the training group was presented, the area under the curve (AUC) for training group was 0.857 (Fig. 7C). In the test group, the signature was validated to have great accuracy with AUC of 0.839, and overall-survival and risk score analysis also

corresponded with the training group (Fig. 8A). The overall-survival, risk score and AUC of the entire set were present in Fig. 8B, the AUC of the entire set was 0.853.

Table 2
Information of ASs included in the signature.

Symbol	As id	Splice type	Exons	From exon	To exon
DECR1	84409	ES	2:3:4.1:4.2	1	5.2
POM121	79925	AP	1	NA	NA
TEX264	65099	AD	1.4	1.3	3
CCNYL1	57200	AT	11	NA	NA
ZBTB7A	46761	AP	2	NA	NA
SLC15A4	25184	AP	5.1	NA	NA
ZBED5	14401	AA	3.1	2	3.2
SIRT3	13599	ES	6.1	5.3	8.1
CREM	11231	AP	13.1	NA	NA
HDGF	8323	AD	1.2	1.1	4
NA represents not available.					

The relationship between molecular subtypes and clinical features.

Based on the PSI value of AS involved in the signature, the samples were classified into 4 molecular clusters (Fig. 9A, B). K-M plot revealed that the difference in the relapse-free survival between cluster 1,3 and cluster 2,4 (Fig. 9C). The association between other clinical features with 4 molecular clusters were shown in Fig. 9D. Base on the AS involved in the signature, the molecular clusters could efficiently distinguish the difference of relapse rate between diverse clusters.

Discussion

The treatment strategy for I–III colon cancer is radical surgery combined with adjuvant chemotherapy, patients can achieve relatively considerable prognosis, however, the relapse after surgery is still a big challenge for colon cancer patients. According to the previous data, approximately 50% of the colon cancer patients would develop relapse after surgery within 2 years[15]. Due to the limitation of the treatment strategy after the relapse, the prognosis of patients who occur relapse is poor. Currently, the better way for treating relapse may be the early detection and the cautious follow-up monitoring for patients with high-risk of relapse. However, the methods utilized to identify the patients who have high-risk to suffer relapse is limited, the prognostic factors for relapse risks mainly contained the number of

examined lymph nodes, mismatch repair (MMR) status, TNM system, and vascular invasion and some other clinicopathological information. Among these risk factors, only the stage of T, the number of examined lymph nodes, and the status of MMR have been examined to have substantial clinical evidence[16, 17].

However, the recurrence of colon cancer is a complex process including many biological dysfunctions, the mechanism of relapse in molecular biology is still unclear and the risk factors that can predict relapse efficiently are awaiting to be elucidated. Alternative splicing is a biological process and plays a vital role in the process of splicing introns to form mature mRNA and protein[18]. The dysregulated ASs can induce the wrong expression of splicing variants or the failure in the expression of right isoforms which indicates its crucial role in diseases. Previous studies also validated the significant role of ASs in colon cancer[19]. However, little literature focused on the role of ASs in the relapse of colon cancer. The profile of aberrantly expressed ASs, the correlation between ASs and SFs, and the interaction between diverse isoforms of proteins which may result in the relapse of colon cancer is urgent to be elucidated.

In the present study, we identified the relapse-associated ASs in I–III colon cancer. 1912 ASs of 1384 mRNA were identified to be aberrantly expressed between relapse group and long-term survival group. Among these 7 subtypes of AS, ES was considered to be the most common one indicating its crucial function in the process of relapse. The proteins involved in the relapse-associated ASs were selected to build a PPI network, the cluster with the highest score of the 30 clusters still contained 789 interactions among 57 proteins. The complexity of the network indicated that relapse was not driven by one single AS-relevant protein, it was a process regulated by the integrated whole system. 10 hub proteins were identified as core proteins in the cluster. Most of them are ubiquitin-related proteins including RBBP6, UBE2D3, UBE3D, HERC4, HERC6, ARIH2, TRIP12, and UBE2D4, the results revealed the ubiquitin was vital in the relapse of colon cancer. Among the 10 hub proteins, RBBP6 owns the most interactions, Xiao et al. found that RBBP6 could promote the metastasis by activating the epithelial-mesenchymal transition (EMT) process[20]. Two-third of relapse was metastasis, our analysis indicated the crucial role of RBBP6 in the process of relapse and previous experimental research also verified our analysis. We also analyzed the correlation between SFs and ASs, DDX39B was the SF that owned the most interaction with ASs. DDX39B was recognized as an efficient alternative splicing activator of exon[21]. Our analysis provided a new view of DDX39B in the regulation of relapse in colon cancer. Among these ASs, ETFB-51319-AP, ETFB-51320-AP, EIF2B1-25097-AT, and EIF2B1-25098-AT, 4 ASs of 2 mRNA owned the most significant role. ETFB was previously reported to be the kind of gene related to host-cell systemic regulation in the development and progression of cancer[22]. EIF2B1 was proved to be a protein compound and is composed of 5 nonidentical subunits, it has an essential role in normal protein production and is indispensable for normal cell function. Human cells were expected to die under two inactivating mutations in one of the EIF2B1 subunits[23]. Our results indicated that ASs of the 2 mRNA may interact with SFs and further cause the relapse of colon cancer. However, the detailed mechanism still needs to be revealed in experiments.

The results of univariate Cox regression analysis identified 986 ASs of 726 mRNA. Many of these ASs and relevant mRNA were reported to have vital functions in diverse cancer including colon cancer, for example, BCLAF1 which had the highest P-value in ES group of our analysis could interact with SF-SRSF10 to regulate the tumorigenic potential of colon cancer[24]. These experimentally validated results prove that our analysis is credible. Studies of some ASs and relevant mRNA are still blank in colon cancer, however, their functions which were previously reported can provide us tips of their role in colon cancer, BTNL8 (highest P-value in AT group) was reported to regulate the T cell biology, indicating it may play a role in immunity therapy of colon cancer[25–27]. After univariate Cox regression analysis, LASSO Cox regression analysis, and multivariate Cox regression analysis with strict screening criteria, we built a robust signature, the AUC of ROC in the training set, test set, and entire set was respectively 0.857, 0.839, and 0.853. The data was almost the same in the three sets indicating that our signature is robust and has similar prediction effects in various data. ASs included in our signature can function as great predictive factors to identify the patients with high-risk of relapse. Most of the ASs involved in the signature were reported to have functions in cancers, many of them also have functions especially in the field of colon cancer. Torrens-Mas et al. found that silence of SIRT3 can improve the efficacy of oxaliplatin[28]. Liu et al. indicated that ZBTB7A can act as a tumor suppressor in the regulation of glycolysis in colon cancer[29]. Finally, we divided the samples into 4 molecular subtypes, and the molecular clusters could efficiently distinguish the difference of relapse rate between diverse clusters, which further validated our signature powerful.

Conclusions

In conclusion, to our knowledge, our study is the first one to provide a profile of relapse-associated ASs in I–III colon cancer and build a robust signature to predict the relapse of I–III colon cancer patients. We further classify the patients into 4 molecular subtypes based on the signature. Our signature may help to better predict relapse of colon cancer.

Declarations

Ethics approval and consent to participate

This is a study from public datasets with minimal risk and we hope to waive the ethics consent.

Consent for publication

We obtained consents to publish this paper from all authors included this study.

Availability of data and materials

The authors declare that the data and materials of this study are available.

Competing interests

Authors declare that they have no competing interests.

Funding

This project was supported by The National Natural Science Foundation of China (81702302 and 81472228)

Author contribution

ZYZ designed and conducted the study. ZYZ wrote the article. JMX, QYF and PZ helped to improve and design the study. YL, YHM, YQX and GDH helped to analyze the data and write the article.

Acknowledgements

We acknowledge Dr. Qingyang Feng for carefully reviewed our paper.

References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA: a cancer journal for clinicians* 2020, **70**(1):7-30.
2. Shi Q, Andre T, Grothey A, Yothers G, Hamilton SR, Bot BM, Haller DG, Van Cutsem E, Twelves C, Benedetti JK *et al*: **Comparison of outcomes after fluorouracil-based adjuvant therapy for stages II and III colon cancer between 1978 to 1995 and 1996 to 2007: evidence of stage migration from the ACCENT database**. *J Clin Oncol* 2013, **31**(29):3656-3663.
3. Becker H: **[Surgery of colorectal carcinoma]**. *Praxis* 1995, **84**(47):1371-1372.
4. Tsai HL, Chu KS, Huang YH, Su YC, Wu JY, Kuo CH, Chen CW, Wang JY: **Predictive factors of early relapse in UICC stage I-III colorectal cancer patients after curative resection**. *J Surg Oncol* 2009, **100**(8):736-743.
5. Van Cutsem E, Nordlinger B, Adam R, Kohne CH, Pozzo C, Poston G, Ychou M, Rougier P: **Towards a pan-European consensus on the treatment of patients with colorectal liver metastases**. *European journal of cancer (Oxford, England : 1990)* 2006, **42**(14):2212-2221.
6. Seo SI, Lim SB, Yoon YS, Kim CW, Yu CS, Kim TW, Kim JH, Kim JC: **Comparison of recurrence patterns between ≤ 5 years and > 5 years after curative operations in colorectal cancer patients**. *J Surg Oncol* 2013, **108**(1):9-13.
7. Kozlovski I, Siegfried Z, Amar-Schwartz A, Karni R: **The role of RNA alternative splicing in regulating cancer metabolism**. *Human genetics* 2017, **136**(9):1113-1127.
8. Nilsen TW, Graveley BR: **Expansion of the eukaryotic proteome by alternative splicing**. *Nature* 2010, **463**(7280):457-463.
9. Climente-Gonzalez H, Porta-Pardo E, Godzik A, Eyras E: **The Functional Impact of Alternative Splicing in Cancer**. *Cell reports* 2017, **20**(9):2215-2226.
10. Oltean S, Bates DO: **Hallmarks of alternative splicing in cancer**. *Oncogene* 2014, **33**(46):5311-5318.

11. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK: **RNA splicing factors as oncoproteins and tumour suppressors.** *Nature reviews Cancer* 2016, **16**(7):413-430.
12. Wan L, Yu W, Shen E, Sun W, Liu Y, Kong J, Wu Y, Han F, Zhang L, Yu T *et al.*: **SRSF6-regulated alternative splicing that promotes tumour progression offers a therapy target for colorectal cancer.** *Gut* 2019, **68**(1):118-129.
13. Sakuma K, Sasaki E, Kimura K, Komori K, Shimizu Y, Yatabe Y, Aoki M: **HNRNPLL, a newly identified colorectal cancer metastasis suppressor, modulates alternative splicing of CD44 during epithelial-mesenchymal transition.** *Gut* 2018, **67**(6):1103-1111.
14. Giulietti M, Piva F, D'Antonio M, D'Onorio De Meo P, Paoletti D, Castrignano T, D'Erchia AM, Picardi E, Zambelli F, Principato G *et al.*: **SpliceAid-F: a database of human splicing factors and their RNA-binding sites.** *Nucleic acids research* 2013, **41**(Database issue):D125-131.
15. Kindler HL, Shulman KL: **Metastatic colorectal cancer.** *Current treatment options in oncology* 2001, **2**(6):459-471.
16. Kelley RK, Venook AP: **Prognostic and predictive markers in stage II colon cancer: is there a role for gene expression profiling?** *Clinical colorectal cancer* 2011, **10**(2):73-80.
17. Tsikitis VL, Larson DW, Huebner M, Lohse CM, Thompson PA: **Predictors of recurrence free survival for patients with stage II and III colon cancer.** *BMC cancer* 2014, **14**:336.
18. Lee SC, Abdel-Wahab O: **Therapeutic targeting of splicing in cancer.** *Nat Med* 2016, **22**(9):976-986.
19. Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, Veitch J, Schweitzer A, Awad T, Sugnet C, Dee S *et al.*: **Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array.** *BMC genomics* 2006, **7**:325.
20. Xiao C, Wu G, Zhou Z, Zhang X, Wang Y, Song G, Ding E, Sun X, Zhong L, Li S *et al.*: **RBBP6, a RING finger-domain E3 ubiquitin ligase, induces epithelial-mesenchymal transition and promotes metastasis of colorectal cancer.** *Cell death & disease* 2019, **10**(11):833.
21. Galarza-Munoz G, Briggs FBS, Evsyukova I, Schott-Lerner G, Kennedy EM, Nyanhete T, Wang L, Bergamaschi L, Widen SG, Tomaras GD *et al.*: **Human Epistatic Interaction Controls IL7R Splicing and Increases Multiple Sclerosis Risk.** *Cell* 2017, **169**(1):72-84.e13.
22. Wan D, Gong Y, Qin W, Zhang P, Li J, Wei L, Zhou X, Li H, Qiu X, Zhong F *et al.*: **Large-scale cDNA transfection screening for genes related to cancer development and progression.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(44):15724-15729.
23. van der Knaap MS, van Berkel CG, Herms J, van Coster R, Baethmann M, Naidu S, Boltshauser E, Willemsen MA, Plecko B, Hoffmann GF *et al.*: **eIF2B-related disorders: antenatal onset and involvement of multiple organs.** *American journal of human genetics* 2003, **73**(5):1199-1207.
24. Zhou X, Li X, Cheng Y, Wu W, Xie Z, Xi Q, Han J, Wu G, Fang J, Feng Y: **BCLAF1 and its splicing regulator SRSF10 regulate the tumorigenic potential of colon cancer cells.** *Nat Commun* 2014, **5**:4581.
25. Vantourout P, Laing A, Woodward MJ, Zlatareva I, Apolonia L, Jones AW, Snijders AP, Malim MH, Hayday AC: **Heteromeric interactions regulate butyrophilin (BTN) and BTN-like molecules governing**

- gammadelta T cell biology.** *Proceedings of the National Academy of Sciences of the United States of America* 2018, **115**(5):1039-1044.
26. Di Marco Barros R, Roberts NA, Dart RJ, Vantourout P, Jandke A, Nussbaumer O, Deban L, Cipolat S, Hart R, Iannitto ML *et al*: **Epithelia Use Butyrophilin-like Molecules to Shape Organ-Specific gammadelta T Cell Compartments.** *Cell* 2016, **167**(1):203-218.e217.
27. Willcox CR, Vantourout P, Salim M, Zlatareva I, Melandri D, Zanardo L, George R, Kjaer S, Jeeves M, Mohammed F *et al*: **Butyrophilin-like 3 Directly Binds a Human Vgamma4(+) T Cell Receptor Using a Modality Distinct from Clonally-Restricted Antigen.** *Immunity* 2019, **51**(5):813-825.e814.
28. Torrens-Mas M, Hernandez-Lopez R, Oliver J, Roca P, Sastre-Serra J: **Sirtuin 3 silencing improves oxaliplatin efficacy through acetylation of MnSOD in colon cancer.** *J Cell Physiol* 2018, **233**(8):6067-6076.
29. Liu XS, Haines JE, Mehanna EK, Genet MD, Ben-Sahra I, Asara JM, Manning BD, Yuan ZM: **ZBTB7A acts as a tumor suppressor through the transcriptional repression of glycolysis.** *Genes & development* 2014, **28**(17):1917-1928.

Figures

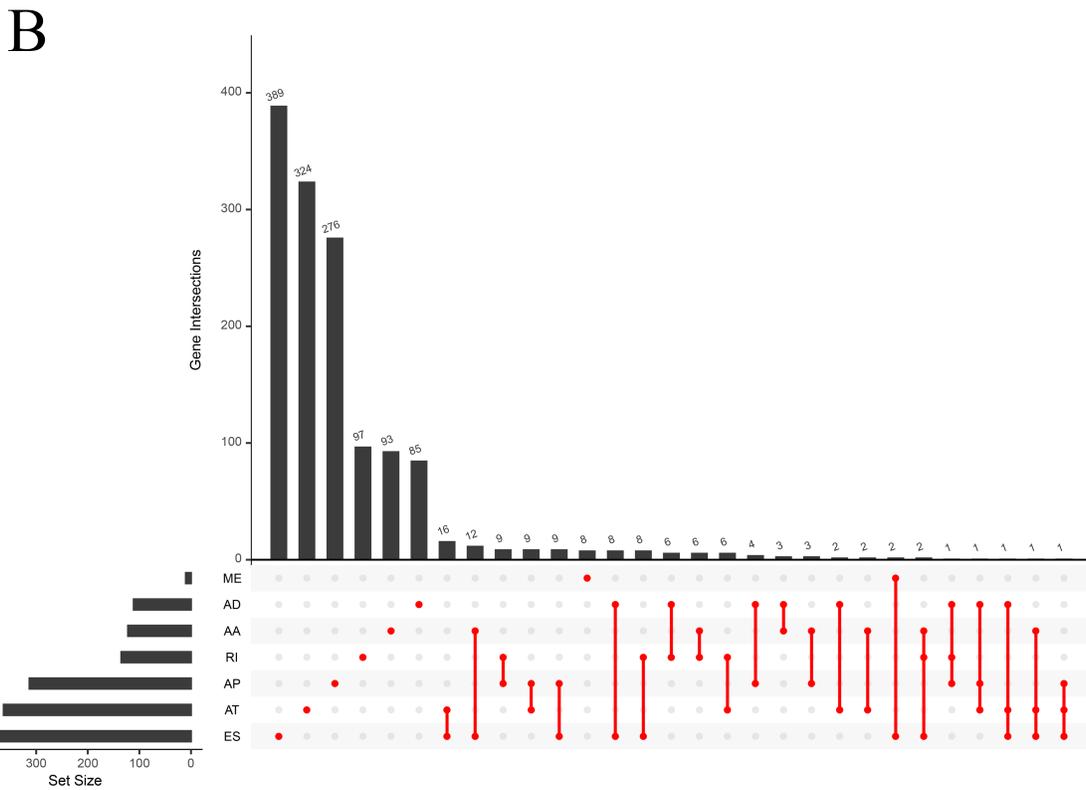
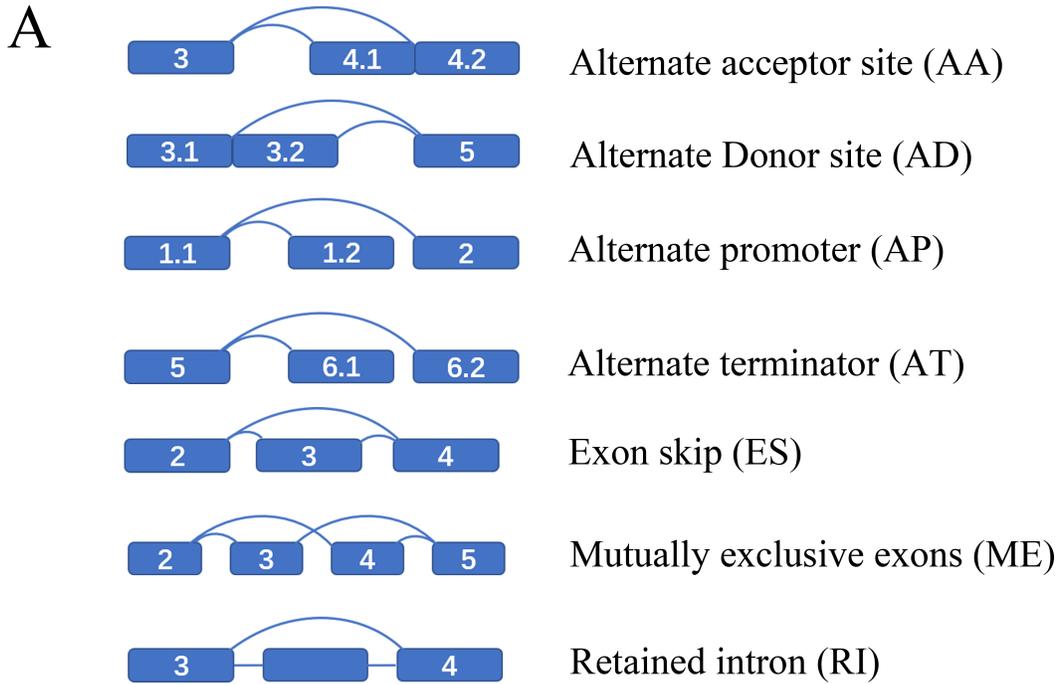


Figure 1

Profile of relapse-associated ASs in colon cancer. (A) Illustration of the 7 subtypes of ASs: Alternate Acceptor site (AA), Alternate Donor site (AD), Alternate Promoter (AP), Alternate Terminator (AT), Exon Skip (ES), Mutually Exclusive Exons (ME), and Retained Intron (RI). (B) Upset plot depicting the overlapping of the 7 subtypes ASs and relevant mRNA. One mRNA owns up to 3 types of ASs.

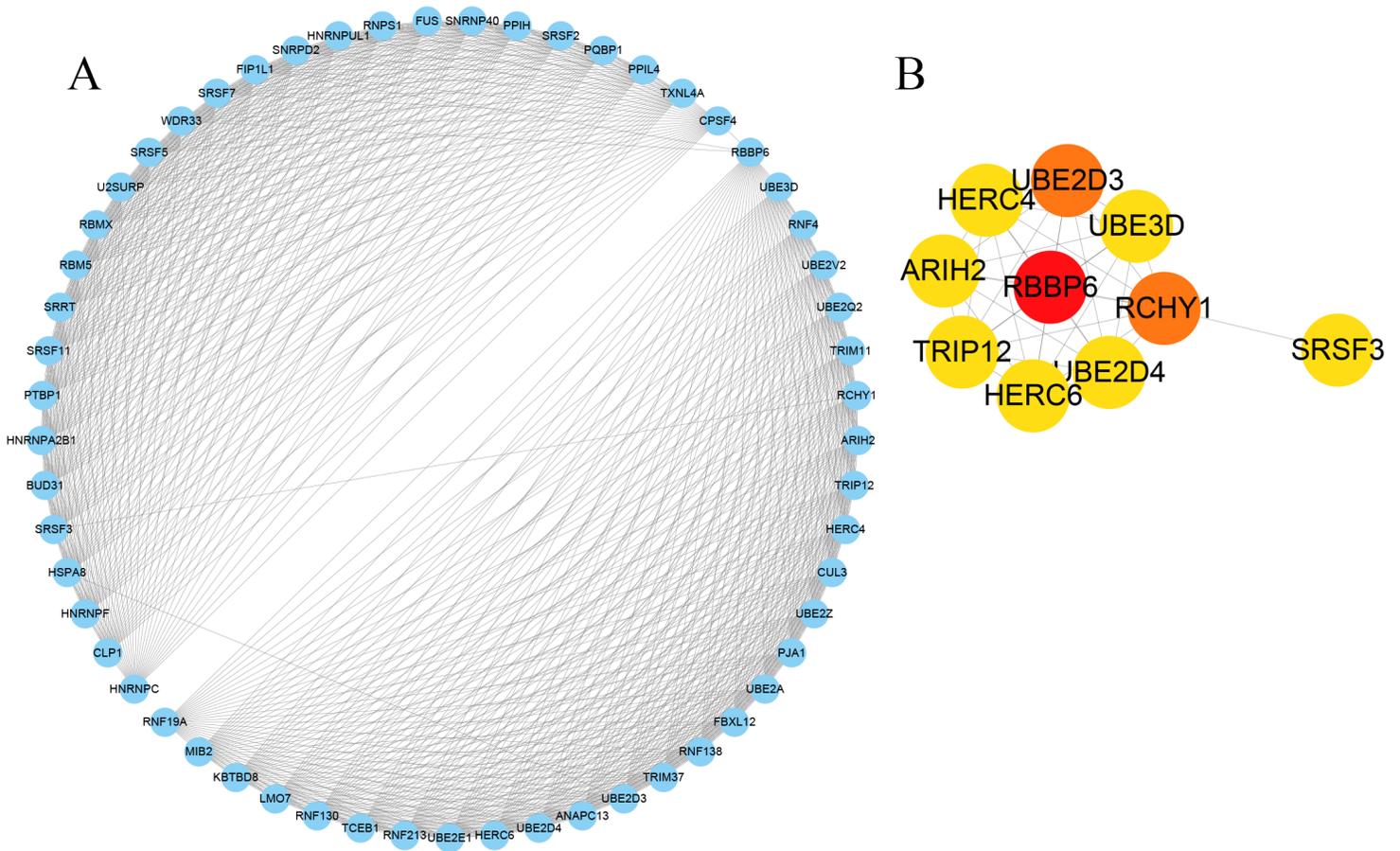


Figure 2

PPI network of genes included in relapse-associated ASs. (A) The cluster that owns the highest score (28.2) among the 30 clusters. (B) Interactions between the 10 hub proteins in the cluster with the highest score. The score of interactions between proteins ranks by color order as follows: red, orange, and yellow.

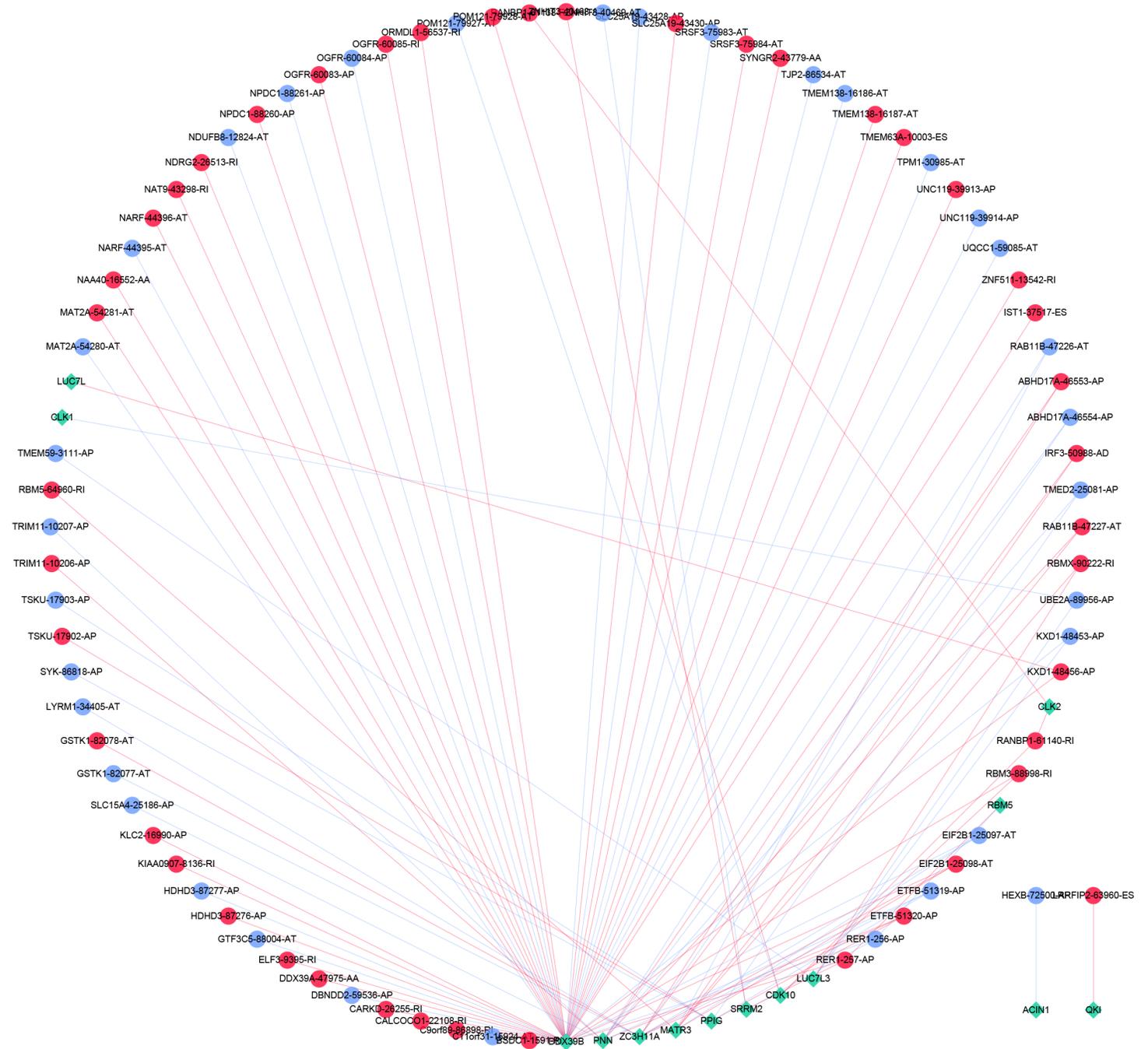


Figure 3

Correlation between relapse-associated ASs and relevant SFs. Shape of the nodes represents the type (circle represents AS, and diamond represents SF). The color of ASs represents the aberrantly expressed type between relapse group and long-term survival group. Red represents upregulation and blue represents downregulation in the relapse group. The color of the line represents the correlation between ASs and SFs. Red represents a positive correlation and blue represents a negative correlation.

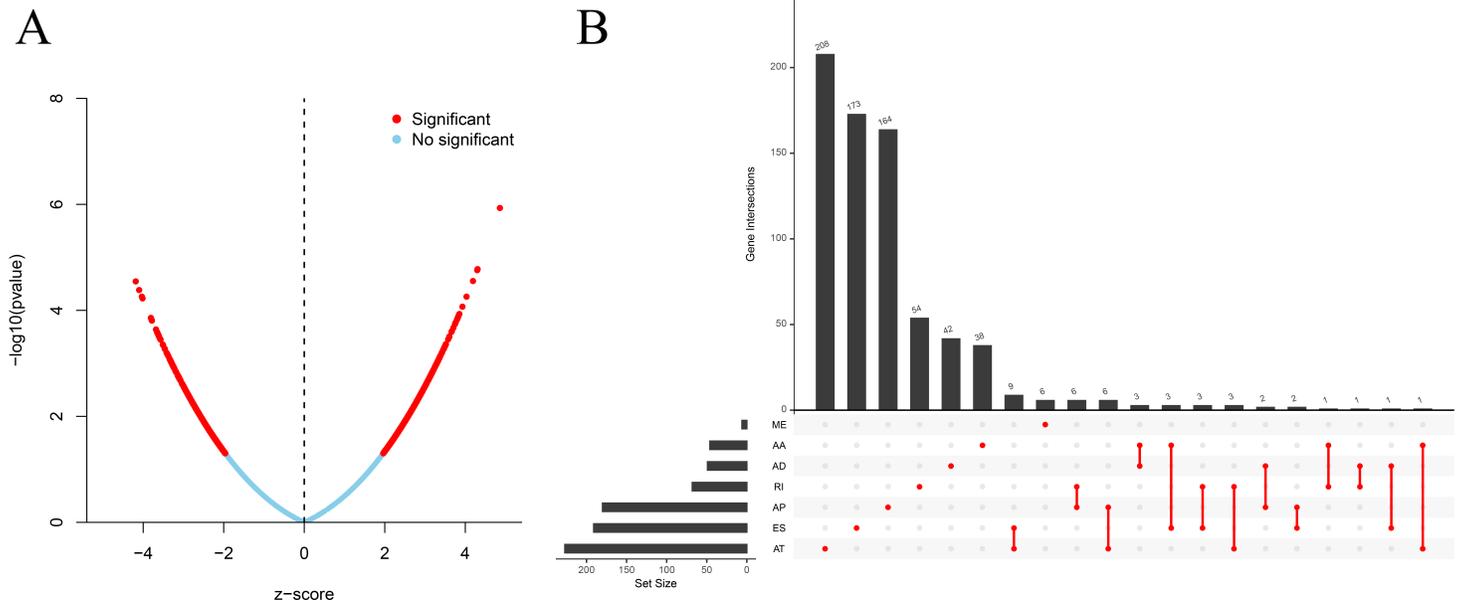


Figure 4

Profile of relapse-associated ASs based on the univariate Regression Cox model. (A) Volcano plot depicting the ratio of statistically significant relapse-associated ASs after univariate Regression Cox analysis. (B) Upset plot depicting the overlapping of the 7 subtypes ASs and relevant mRNA. One mRNA owns up to 2 types of ASs.

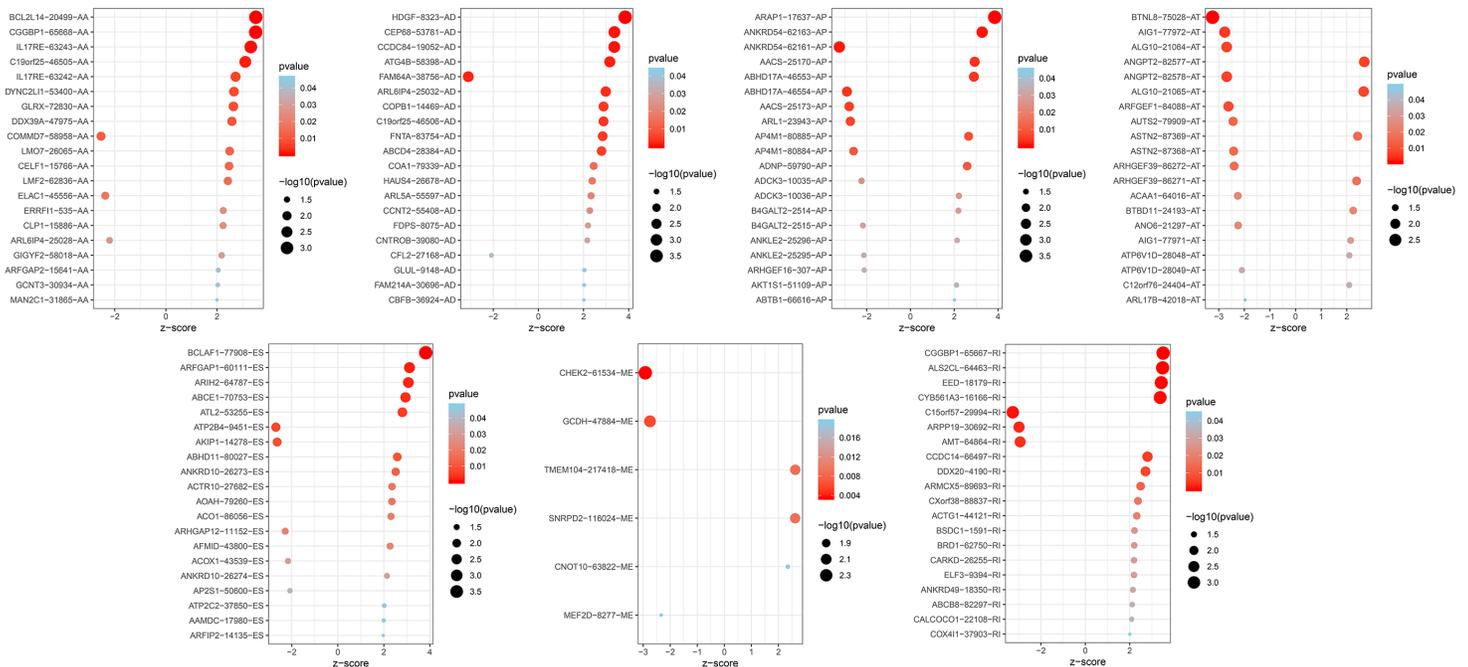


Figure 5

Bubble plot depicting the distribution of 7 types of relapse-associated ASs after univariate Regression Cox analysis.

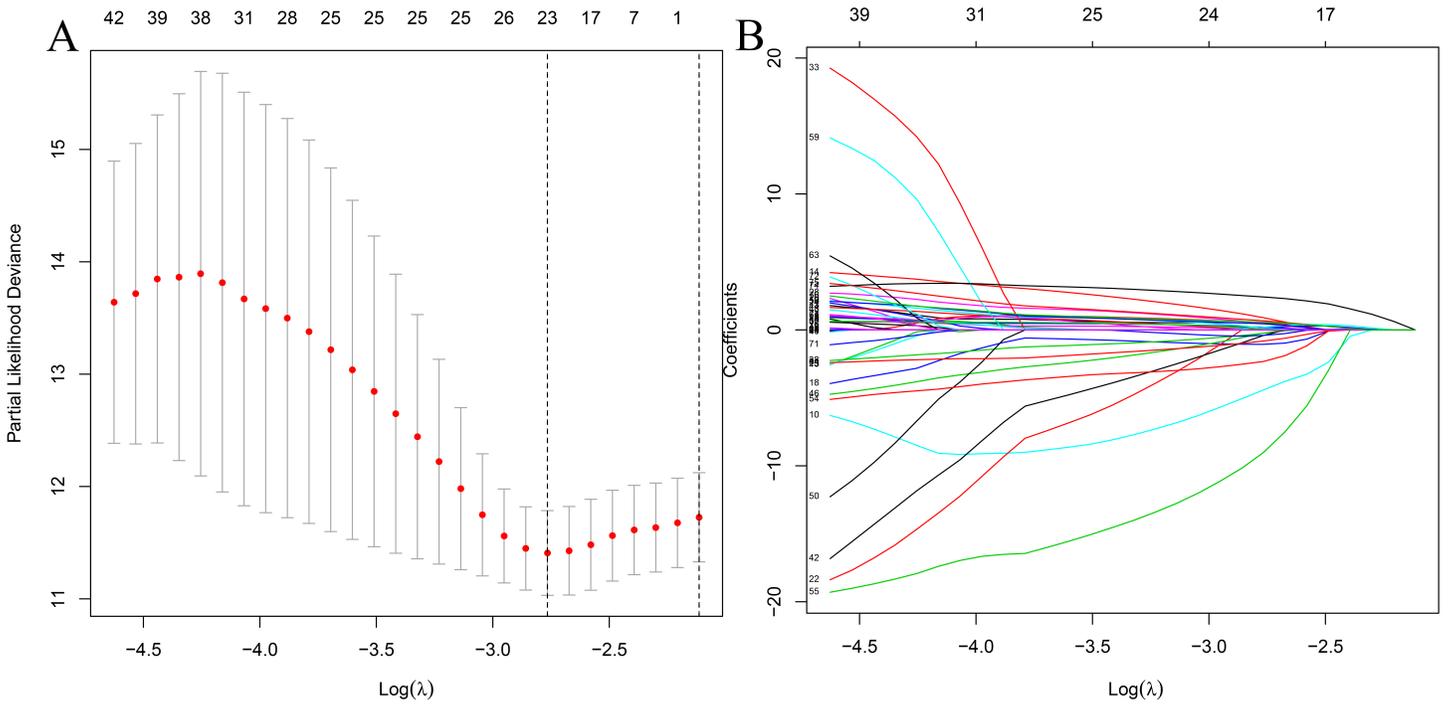


Figure 6

Results of LASSO Cox regression analysis. (A) The correlation between $\text{log}(l)$ and the deviance. (B) The correlation between $\text{log}(l)$ and the coefficients of relevant ASs.

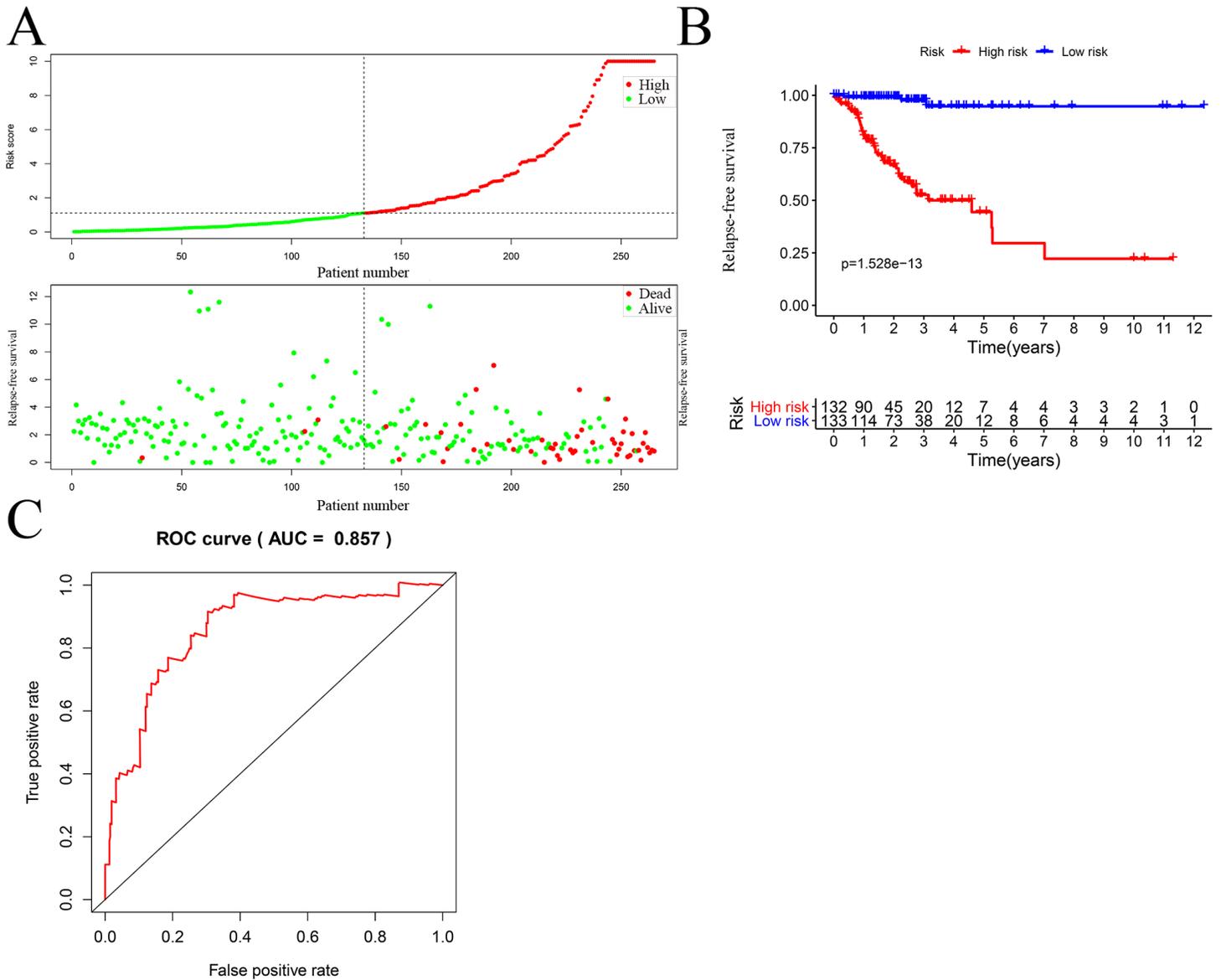


Figure 7

Prognostic value of signature in the prediction of the relapse in colon cancer in the training group. (A) The risk score analysis including the risk scores of patients (upper) and survival status of patients (down). (B) The survival analysis of relapse in colon cancer with the relevant number of exposed to risk. (C) ROC curve.

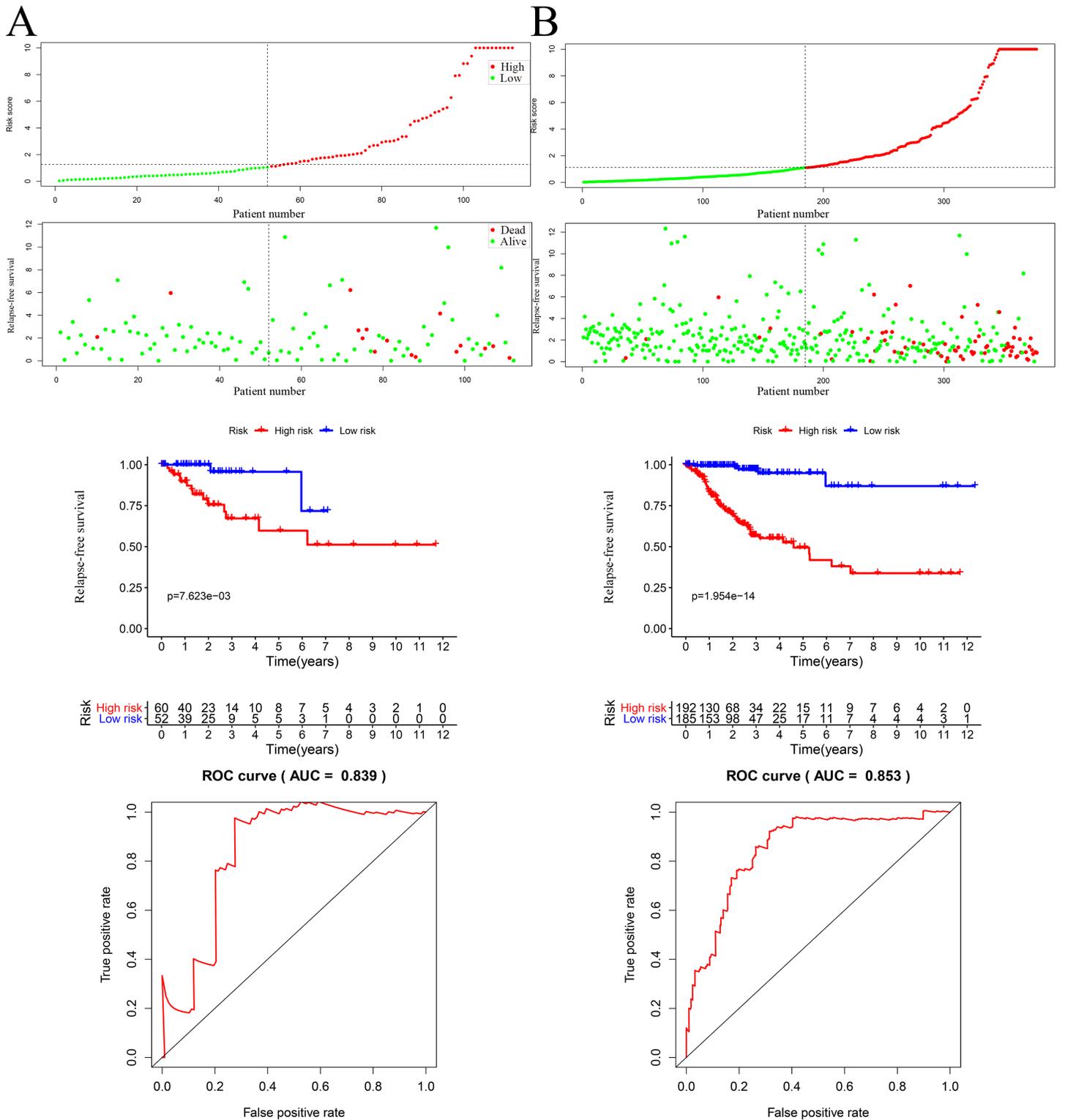


Figure 8

Validation of the prognostic value of signature in the prediction of the relapse in colon cancer. (A) The test group. (B) The entire group. The upper panels are the risk score analysis including the risk scores of patients (upper) and survival status of patients (down). The middle panels are the survival analysis of relapse in colon cancer with the relevant number of exposed to risk. The bottom panels are the ROC curves.

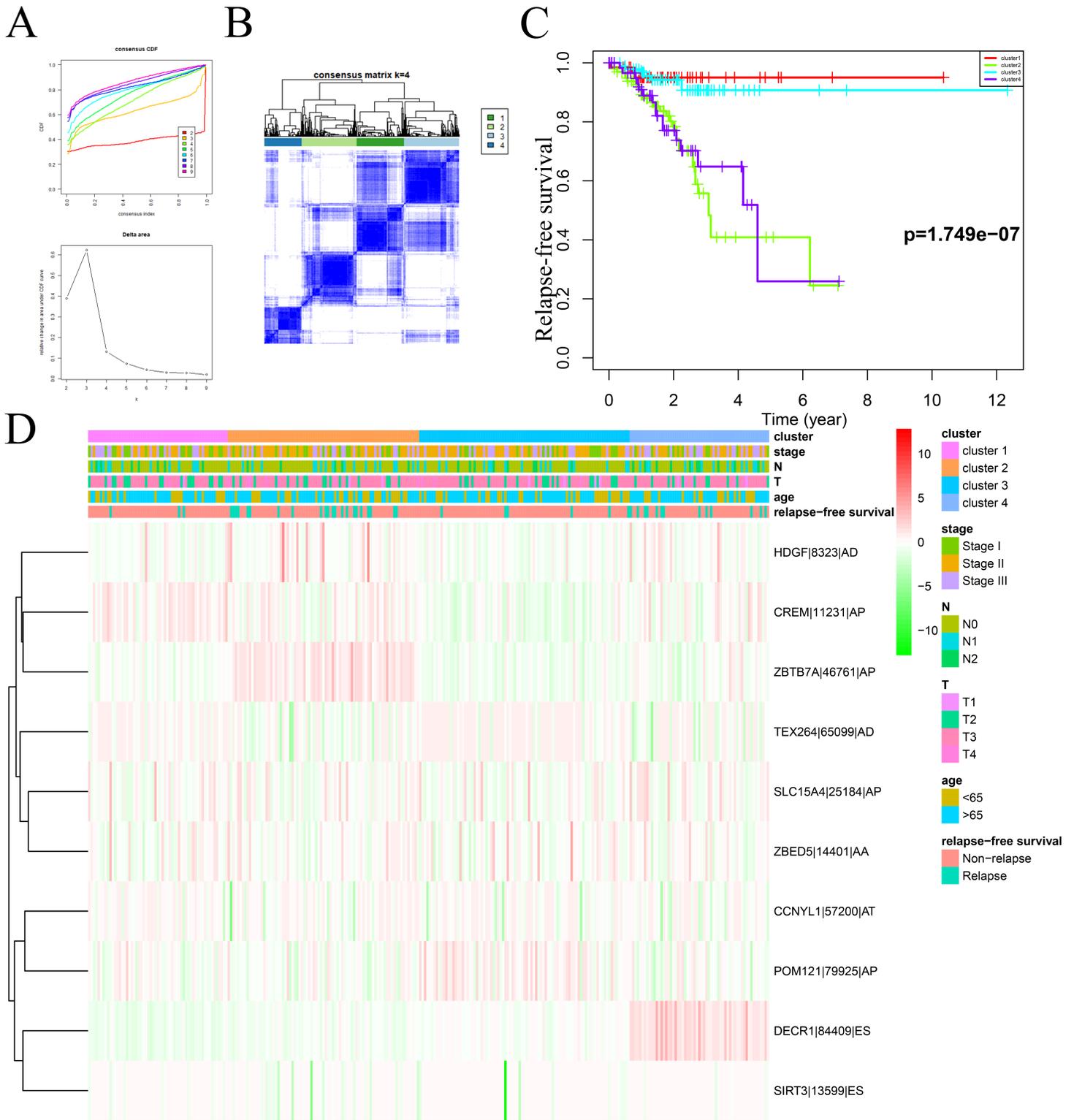


Figure 9

Relapse-associated AS clusters related to prognosis and molecular subtypes. (A) Elbow and Gap analysis for different numbers of clusters ($k = 2$ to 9). (B) Consensus heatmap depicted 4 clusters of samples. (C) Kaplan-Meier survival analysis of patients with different clusters on relapse-free survival. Depicted P-values are from log-rank tests. (D) Spine plots of relationship between 4 molecular subtypes and clinical features.