



Identification of Novel Alternative Splicing Events Associated With Tumorigenesis, Protein Modification, and Immune Microenvironment in Early-Onset Gastric Cancer

Jian Zhang¹, Ajay Goel² and Lin Zhu^{1*}

¹ Department of Pharmaceutical Sciences, Irma Lerma Rangel College of Pharmacy, Texas A&M University, College Station, TX, United States, ² Beckman Research Institute, City of Hope Comprehensive Cancer Center, Biomedical Research Center, Monrovia, CA, United States

OPEN ACCESS

Edited by:

Jia Wei,
Nanjing Drum Tower Hospital, China

Reviewed by:

Riyaz Mir,
All India Institute of Medical Sciences,
India
Debleena Ray,
Duke-NUS Medical School, Singapore

*Correspondence:

Lin Zhu
lzhu@tamu.edu

Specialty section:

This article was submitted to
Gastrointestinal Cancers,
a section of the journal
Frontiers in Oncology

Received: 11 December 2020

Accepted: 10 May 2021

Published: 08 June 2021

Citation:

Zhang J, Goel A and Zhu L (2021)
Identification of Novel Alternative
Splicing Events Associated With
Tumorigenesis, Protein Modification,
and Immune Microenvironment in
Early-Onset Gastric Cancer.
Front. Oncol. 11:640272.
doi: 10.3389/fonc.2021.640272

Background: Alternative splicing (AS), e.g. the tandem alternative polyadenylation (TAPA), has emerged as major post-transcriptional modification events in human disease. However, the roles of the AS and TAPA in early-onset gastric cancer (EOGC) have not been revealed.

Methods: The global AS profiles of 80 EOGC patients were analyzed. The EOGC-specific AS events (ESASs) were identified in both the EOGC and adjacent non-tumor tissues. The functional enrichment analysis, Splicing network, Alternative Polyadenylation (APA) core factor network, and cell abundancy analysis were performed. Furthermore, the landscapes of the AS events in the varied subtypes of the EOGC patients were evaluated.

Results: Overall, 66,075 AS events and 267 ESASs were identified in the EOGC. Furthermore, 4809 genes and 6152 gene isoforms were found to be aberrantly expressed in the EOGC. The Gene Ontology (GO) and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analyses showed that the significant pathway alterations might exist in these AS events, genes, and gene isoforms. Moreover, the Protein-protein interaction (PPI) network analysis revealed that the UBC, NEK2, EPHB2, and DCTN1 genes were the hub genes in the AS events in the EOGC. The immune cell infiltration analysis indicated a correlation between the AS events and the cancer immune microenvironment. The distribution of the AS events in varied EOGC subtypes, protein phosphorylation and glycosylation was uneven.

Conclusion: The study highlighted the vital roles of the AS in the EOGC, including modulating the specific protein modification and reshaping the cancer immune microenvironment, and yielded new insights into the diagnosis of the EOGC as well as cancer treatment.

Keywords: alternative splicing, immune microenvironment, early-onset gastric cancer, protein phosphorylation, protein glycosylation, alternative polyadenylation

INTRODUCTION

Gastric cancer (GC), a morbid and frequently lethal malignancy, is one of the most common cancers and the leading cause of cancer death worldwide, especially in East Asia (1). For most patients, the GC is usually associated with the unfavorable prognoses and can be only diagnosed at the relatively late stages, resulting in the limited treatment options. The major types of cancer, including the brain, cervix, esophageal squamous cell carcinoma, Kaposi sarcoma, larynx, lung, and non-Hodgkin lymphoma, remain a relatively low incidence rate among young adults. However, the cases of gastric non-cardia cancer in young adults kept increasing from 1995 to 2014 in the US (2). The GC occurring in the patients under the age of 45 is defined as EOGC (3). Though the genetic and environmental factors have been identified to be associated with the GC, the occurrence of EOGC remains largely unexplained.

Studies have profiled the alternative splicing (AS) events in The Cancer Genome Atlas (TCGA) gastric carcinomas (4), Epstein-Barr virus (EBV) associated gastric carcinomas (5), and gastric cell lines (6, 7), which ascertained the important roles of the AS events in the GC. Other accumulating evidence also show that the somatic CDH1 or TGFBR1 gene mutation and proteogenomic alteration are remarkable and may unleash the GC in younger adults (8, 9), suggesting the importance of the post-transcriptional regulation in the EOGC. The AS, one of the key post-transcriptional events, can generate various mRNA transcripts (isoforms) and affect their stability as well. As a result, the downstream protein variants translated from these mRNA isoforms vary in their sequences (10). On the one hand, the protein synthesis is directed by the human genome, but on the other hand, the protein diversity is ensured to satisfy various biological processes (11). Li et al. found the novel AS events and peptides during the mouse stomach formation (12). Furthermore, a comparative genomic study has been performed and identified a distinct molecular expression profile in the EOGC rather than the late onset gastric cancer (LOGC) (13). All these studies suggest that EOGC may have a varied AS event landscape compared to other types of GC.

The TAPA event introduces two or more poly(A) sites to the 3' UTR of a gene which is distributed across a wide variety of cancers and modulates the sensitivity of certain anticancer drugs (14, 15). The shift of the TAPA events shows a cancer-specific and cell organelle-specific mode (16). It was reported that the APA-guided shortening of the NET1 gene 3' UTR enhanced the mRNA transcriptional activity and promoted the metastasis of gastric cancer cells (17). Some studies defined the TAPA as a subtype of AS, while others consider it a pre-mRNA processing (18). In this study, the TAPA was analyzed as the AS.

So far, only a few genome-wide association studies on the EOGC have been performed and none of them have taken into account the effects of the AS events on the EOGC. To our knowledge, the role of the AS in the EOGC remains vague. Herein, the major aims of this study are to analyze the AS events including the TAPA events and to evaluate their roles in the EOGC. The study contains the large-scale RNA sequencing data generated from the EOGC samples. The human transcriptome is

surveyed to identify the genome-wide AS events in the EOGC and the ESAS are identified thereafter. Furthermore, the biological functions of these events are explored. The AS events in different EOGC subtypes are elucidated and their association with the molecular features and tumor immune microenvironment are also profiled.

METHODS

Data Source

The RNA-Seq data of an independent cohort (tumor tissues and adjacent normal tissues from 80 EOGC patients) was retrieved from the European Nucleotide Archive (study accession: PRJNA508414). One adjacent normal tissue sample (SRR8281377) was excluded from this dataset due to the file error when the Whippet software read and processed the data. These patients were histologically diagnosed as GC and with the age at the time of surgery ≤ 45 years.

RNA-Seq Analysis of the AS Events

Paired-end reads per sample data were generated using the HiSeq 2500 platform sequencer (Illumina). The fastq file was processed by Whippet (version 0.11) on our Linux system (19) and aligned to the Homo sapiens GRCh37.75 genome assembly (ftp://ftp.ensembl.org/pub/release-75/fasta/homo_sapiens/dna/Homo_sapiens.GRCh37.75.dna.primary_assembly.fa.gz). Besides default settings, Whippet was run with the "biascorrect" option to implement the GC-content and 5' sequence bias correction. The Percent Spliced-in Index (PSI or ψ) value for the AS events as well as read count of genes and gene isoforms were generated by Whippet. The output file (diff. format) for the comparative analysis of PSI was generated by Whippet-delta.jl using default parameters. The comparative analysis of genes and gene isoforms was computed by the limma (20) and/or edgeR (21) package.

The PSI value (ranging from 0.0 to 1.0) was defined as the proportional abundance of certain AS events and was calculated for the eight types of the AS events. The AS events were determined using the maximum likelihood estimation by the expectation-maximization (EM) algorithm. To generate more accurate AS event profiling, the stringent filters (percentage of samples with PSI values $> 95\%$, minimum PSI value > 0.05) were implemented. For the gene and gene isoform profiles, genes or gene isoforms with the minimum Counts Per Million (CPM) > 0.5 and percentage of samples with the read count values $> 95\%$ were included.

Identification and Pathway Analysis of the ESAS, Genes, and Gene Isoforms

To filter the ESAS in the gastric cancer, the PSI values of the AS events were compared between the tumor and the matched adjacent normal tissues. The AS events with the missing data exceeding 90% of all subjects were excluded. Statistical analysis was performed in the differentially expressed AS events with > 0.9 probability and absolute log₂ fold change

≥ 1 . The genes or gene isoforms with the absolute \log_2 fold change ≥ 1 and false discovery rate (FDR) < 0.05 were defined as the EOGC-specific genes or gene isoforms. Interactive sets among the eight types of the AS were visualized by UpSetR (version 1.4.4) (22).

The associations between the parent genes of the ESAS and biological annotation terms (Gene Ontology [GO] and Kyoto Encyclopedia of Gene and Genome [KEGG] pathway) were detected using the clusterProfiler package (3.16.0) (23). The NetworkAnalyst updated on 05/2020 was applied to analyze the parent genes of each ESAS for constructing the visualized PPI network (24). The STRING Interactome was selected in the PPI database (confidence score cut-off value 900).

Immune Cell Infiltration and ESAS Event Analysis

The differential gene expression between the tumor and normal tissues and the percentage of the cell abundance in the TCGA STAD cohort were calculated by TIMER 2.0 (25). The gene expression data from our study was input into ImmuCellAI (26) to impute the abundance of 24 types of immune cells in the EOGC. Spearman's rank correlation analysis was used to calculate the correlations between the PSI values of the ESAS and immune cell type. The threshold of Spearman's rank correlation coefficients was set to > 0.4 or < -0.4 , and the BH adjusted p-value < 0.05 .

Construction of Correlation Network Among the Splicing Factors, APA Core Factors, and ESAS Events

From the database (27) and publication (28), we selected 71 experimentally validated human splicing factors and 22 APA core factors to build the splicing factor correlation network and APA core factor correlation network. Spearman's rank correlation analysis was used to impute the correlations between the expression of splicing factors or APA core factors and PSI values of ESASs or TAPAs. The threshold of Spearman's rank correlation coefficients was set to > 0.4 or < -0.4 , and the BH adjusted p-value < 0.05 . The correlation network was constructed and visualized by Cytoscape (version 3.7) (29). All statistical analyses were performed by the R language (30) and p-value < 0.05 was considered statistically significant unless specified.

Identification of the EOGC Subtype Relevant AS Events

The data of the EOGC EBV status, microsatellite instability (MSI) status, protein phosphorylation subtype, and protein glycosylation subtype were accessed from Mun et al. (9). The AS events which have > 0.9 probability of being differentially expressed among the different EBV status, MSI status, protein phosphorylation subtypes, and protein glycosylation subtypes were defined as the subtype relevant AS events in the EOGC. The groups with the sample numbers of less than four were not included in the analysis. The PSI difference among the groups was required to be larger or equal 0.1 for at least two groups to

ensuring the biological significance. The analysis was performed for each subtype separately.

RESULTS

Overview of the AS Events in the EOGC Cohort

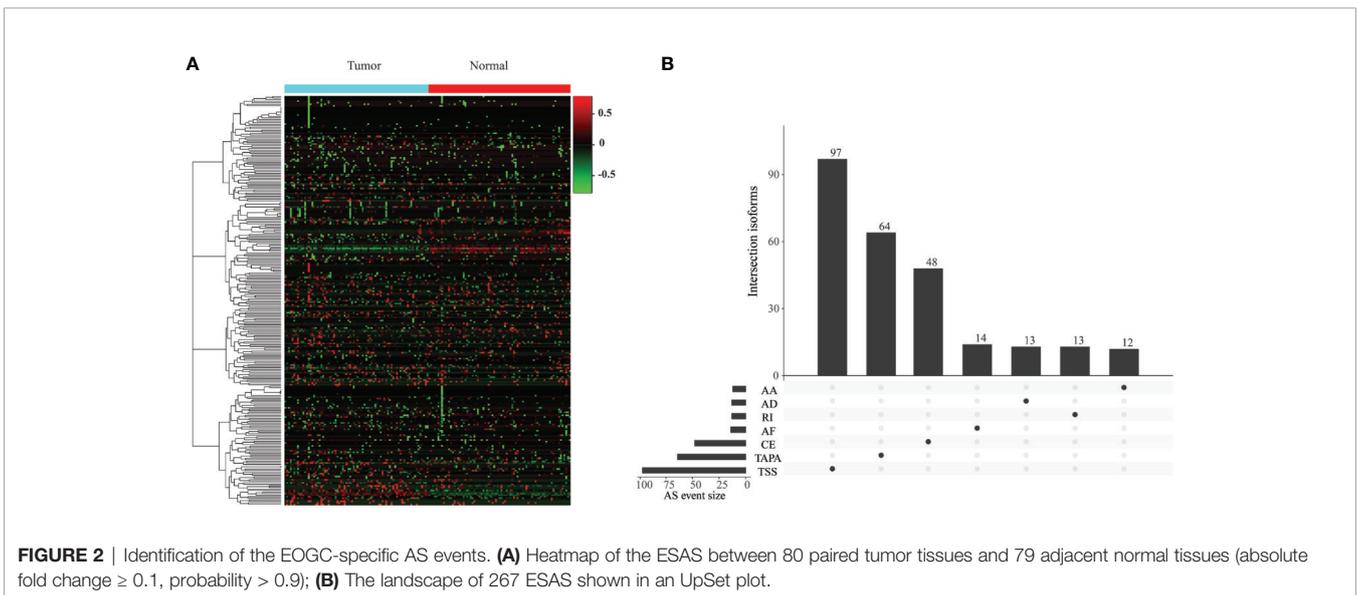
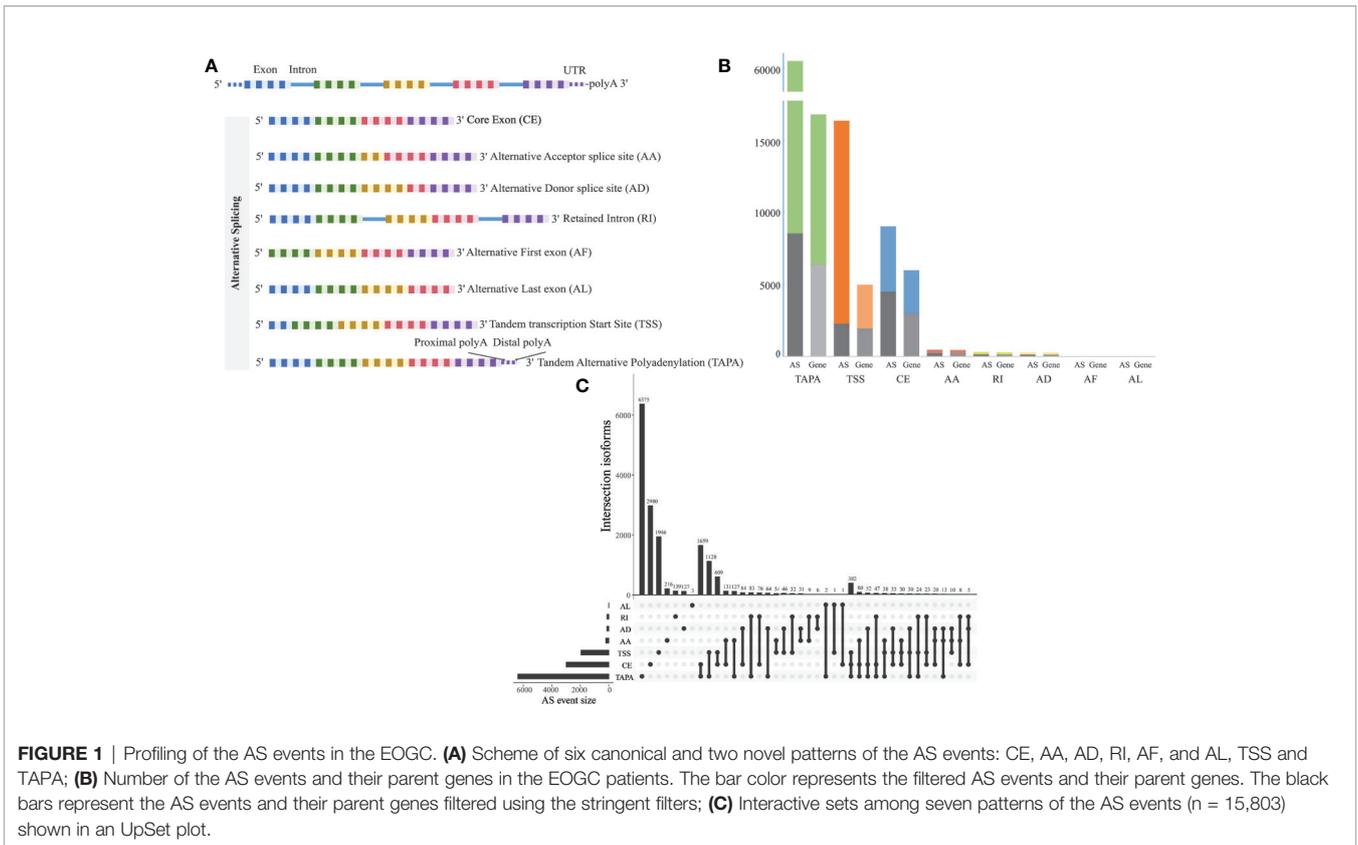
The corresponding RNA-Seq data of 80 EOGC patients were used to establish the integrated AS event profiling. We identified 66,075 AS events from 11,282 genes, which accounted for 53.0% of non-redundant human protein-coding genes (31). Beside two special patterns of the AS events including the tandem transcription start site (TSS) and tandem alternative polyadenylation site (TAPA), these AS events were classified into six canonical splicing patterns: core exon (CE), alternative acceptor splice site (AA), alternative donor splice site (AD), retained intron (RI), alternative first exon (AF), and alternative last exon (AL), as illustrated in **Figure 1A**. Among these splicing patterns, the TAPA occurred most frequently (71.0%) (**Figure 1B**).

Some gene isoforms showed very low redundancy, thus, we screened the AS events with a series of filters (percentage of samples with PSI values $\geq 75\%$, PSI value ≥ 0.05). Consequently, a total of 15,803 AS events from 8,359 genes were obtained. After filtering, the AF events were excluded. The analysis indicated that the TAPA was still the top AS pattern (54.0%) (**Figure 1C**). After removal of the duplicated genes in one AS pattern, the Upset plots were created to quantitatively analyze the interactive sets of the remaining seven patterns of the AS events. As shown in **Figure 1C**, most of the genes had more than one pattern of the AS events, among which some genes had up to three different splicing patterns.

Identification of the ESAS Events

To identify the EOGC-specific AS events, we compared the PSI values between 80 paired tumor tissues and 79 adjacent normal tissues. A total of 267 EOGC-specific AS events (ESASs) from 228 genes were identified (**Figure 2A** and **Table S1**). Although a large number of the AS events were detected in the EOGC cohort, a relatively small proportion of the AS events were identified as the ESAS. The TSS and TAPA events accounted for 36.3% and 24.0% of the ESASs, respectively (**Figure 2B**). The uneven distribution of the AS patterns in the tumor and adjacent normal tissues indicated that they might play important roles in the early-onset gastric tumorigenesis.

For the single gene with multiple ESAS events, the regulatory directions (up- or down-regulation) of varied ESAS events between the tumor and normal tissues can be either the same or opposite. Eight genes (COL5A1, CBWD1, SLC47A2, SSPO, AL020989.1, RPL5P1, AACSP1, and AC022905.1) exhibited the same regulatory direction of AS events; while the other 22 genes showed the opposite regulatory direction (**Table S2**). Interestingly, the proportions of certain AS patterns between the ESAS and entire AS events were inconsistent. The TAPA, a top pattern among all AS events (54.0%), only contributed to 24.0% of the ESAS events.



Profile of the EOGC-Specific Genes and Gene Isoforms

Based on the above results, we further studied how the dysregulated AS events affected the expressions of the genes and gene isoforms. We identified 4809 genes from a total of

16383 genes, and also identified 6152 gene isoforms from a total of 16283 gene isoforms (CPM > 0.5 and the percentage of samples with read count values $> 95\%$).

To generate EOGC-specific genes and gene isoforms, we compared the read count value of the genes or gene isoforms

from the tumor and normal or adjacent non-tumor tissues by the limma-voom (32). After refining, 373 genes and 469 gene isoforms were identified as the EOGC-specific genes and gene isoforms, respectively (absolute log₂ fold change ≥ 1 and FDR < 0.05 , **Figures 3A, B** and **Table S3**). In addition, we profiled the enriched TF binding motifs in the promoters of the EOGC-specific genes and gene isoforms (**Table S4**). A total of four genes and three gene isoforms with ESAs were differentially expressed in the EOGC (**Figure 3C**).

Pathway and Protein-Protein Interaction (PPI) Network of the ESAS Events and EOGC-Specific Genes and Gene Isoforms

To understand the biological roles of the ESAS and EOGC-specific genes and gene isoforms, the GO and KEGG pathway analyses were performed. The results showed that the ESAS parent genes were linked to the GO terms of carboxylic acid biosynthetic process (GO:0046394), organic acid biosynthetic process (GO:0016053), purine-containing compound metabolic process (GO:0072521), collagen-containing extracellular matrix (GO:0062023), etc. These ESAS parent genes were also found to be enriched in the KEGG signaling pathways, including Nicotine addiction (hsa05033) and Arginine and proline metabolism

(hsa00330) (**Figure 4A** and **Table S5**). The GO and KEGG pathway results of the EOGC-specific genes and gene isoforms were also included in the supplementary table (**Table S5**).

The pathway analysis results showed that the parent genes of ESAS might play a vital role in regulating the cancer-related biological processes. The parent genes of the ESAS were further analyzed by the PPI network. Using the Zero-order network, we built the PPI network of the ESAS parent genes, which contained 15 nodes, 14 edges and 15 seeds (**Figure 4B** and **Table S6**). We found that the UBC was the hub gene in the network. Meanwhile, based on the first-order network, we also built 29 and 27 PPI networks using the EOGC-specific genes and gene isoforms, respectively (**Figures 4C, D** and **Table S6**). The UBC, NEK2, EPHB2, and DCTN1 genes were identified as the hub genes in two or more of these networks (DCTN1 was identified as hub gene in the section of protein modification analysis).

Relationship Between the ESAS and Immune Cell Infiltration

To untangle the relationship between the hub genes and gastric tumorigenesis, we analyzed the hub genes in TCGA STAD dataset using the TIMER 2.0 database. We found that the UBC, NEK2, EPHB2, and DCTN1 genes were dysregulated in

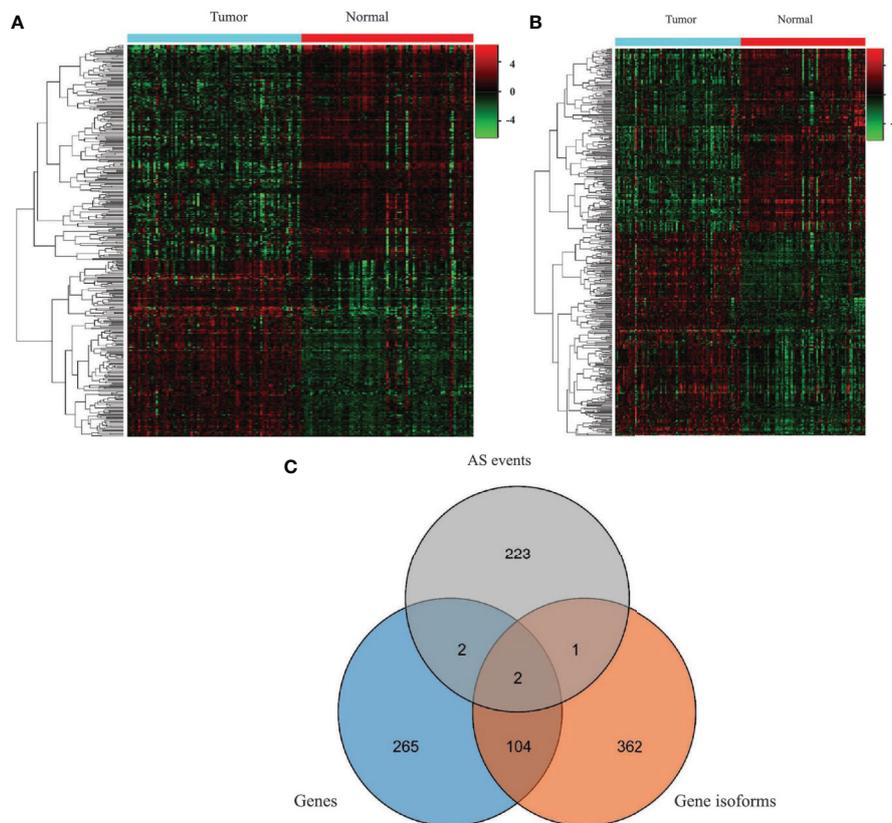


FIGURE 3 | Identification of the EOGC-specific genes and gene isoforms. **(A)** Heatmap of the EOGC-specific genes between 80 paired tumor tissues and 79 adjacent normal tissues (absolute fold change ≥ 1 , FDR < 0.05); **(B)** Heatmap of the EOGC-specific gene isoforms between 80 paired tumor tissues and 79 adjacent normal tissues (absolute fold change ≥ 1 , FDR < 0.05); **(C)** Illustration of the intersection set of the ESAS, EOGC-specific genes, and gene isoforms by the Venn diagram.

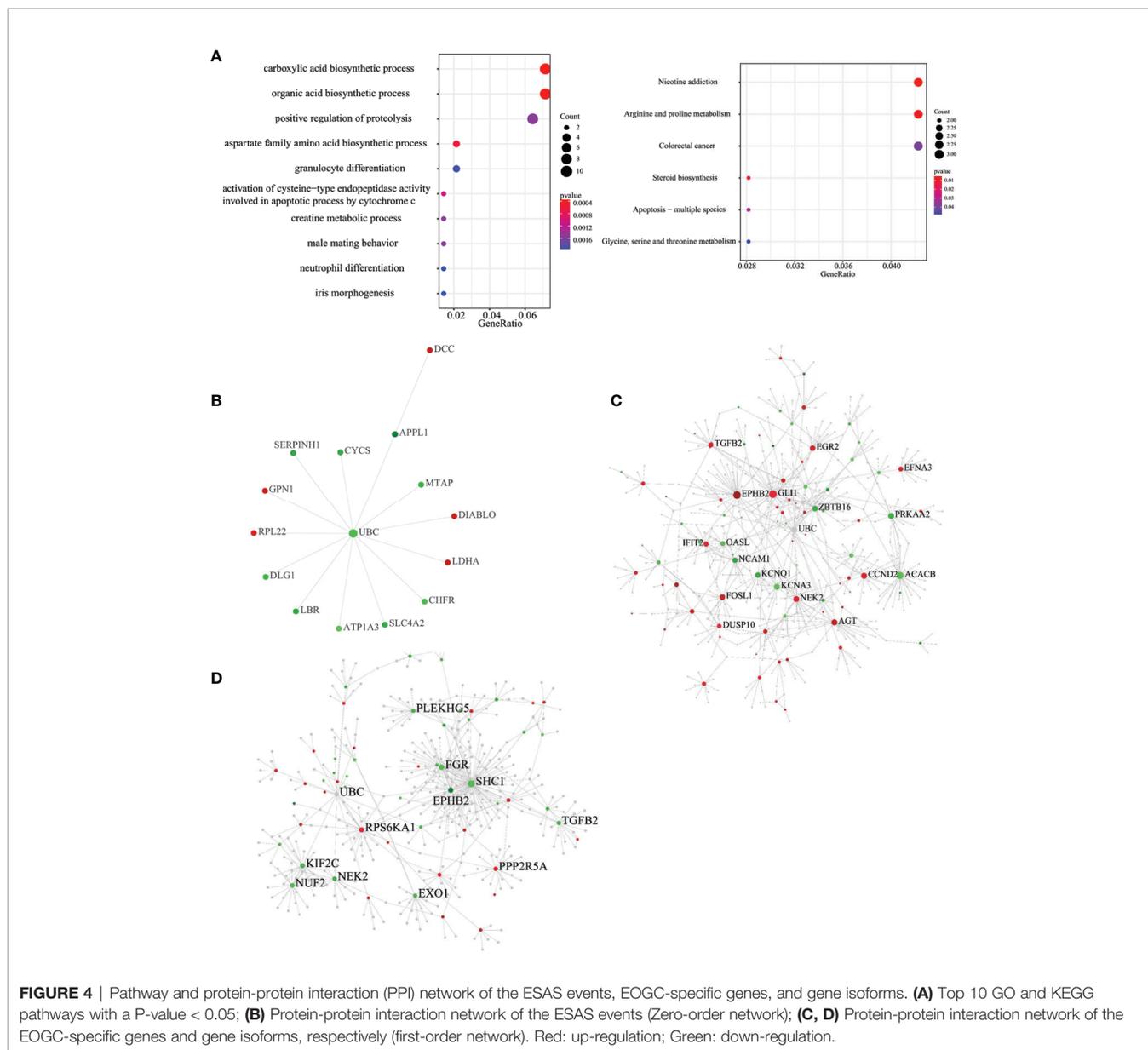
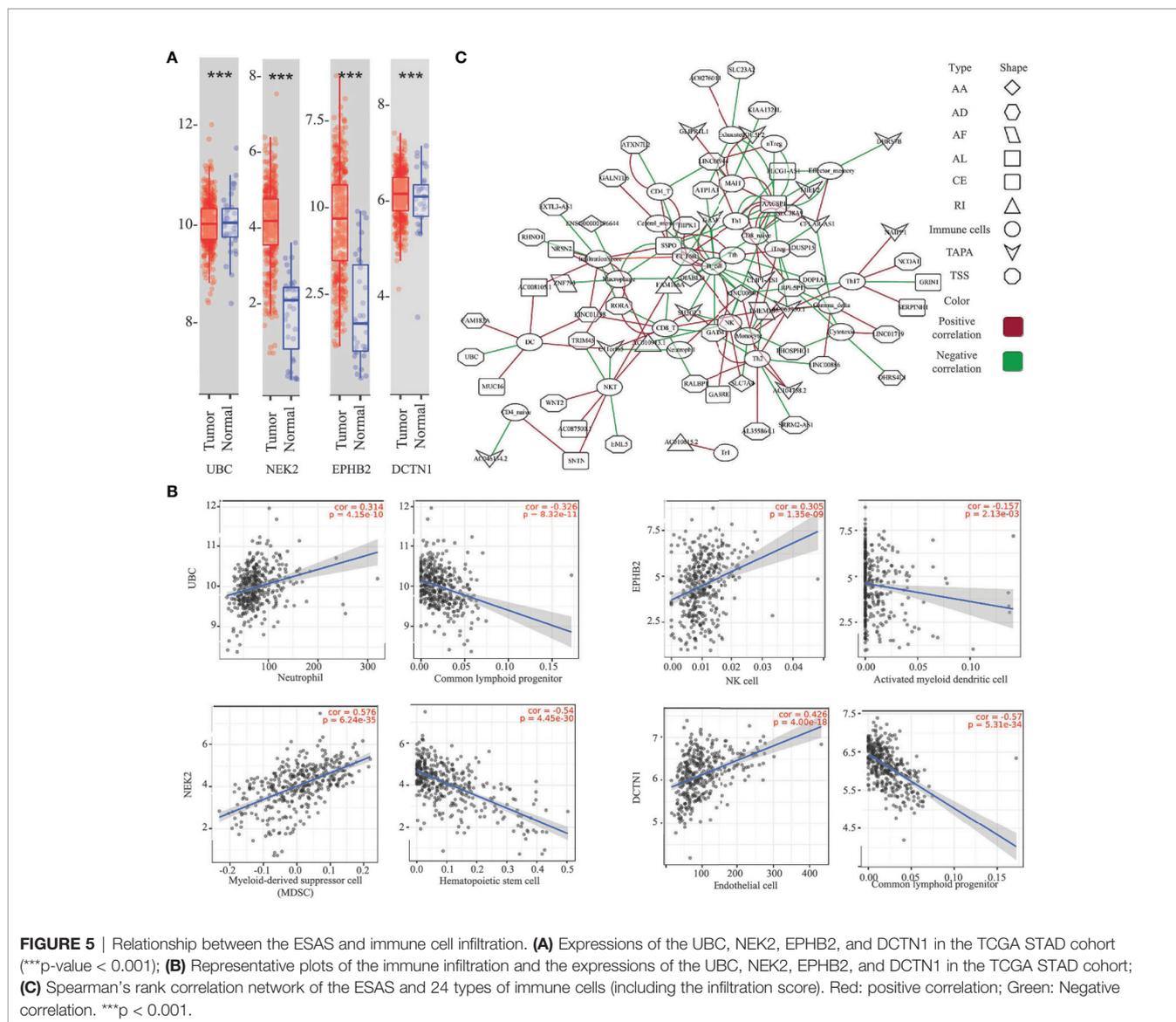


FIGURE 4 | Pathway and protein-protein interaction (PPI) network of the ESAS events, EOGC-specific genes, and gene isoforms. **(A)** Top 10 GO and KEGG pathways with a P-value < 0.05; **(B)** Protein-protein interaction network of the ESAS events (Zero-order network); **(C, D)** Protein-protein interaction network of the EOGC-specific genes and gene isoforms, respectively (first-order network). Red: up-regulation; Green: down-regulation.

the TCGA STAD dataset (p-value < 0.001, **Figure 5A**). The correlation between the expression of the hub genes and cell abundances in the STAD dataset was analyzed. We found that the expression of the UBC gene was positively correlated with the infiltration level of common lymphoid progenitor cells but was negatively correlated with the infiltration level of neutrophil cells. The expression of the NEK2 gene was positively correlated with the infiltration level of myeloid-derived suppressor cells (MDSC) but was negatively correlated with the infiltration level of hematopoietic stem cells. The expression of the EPHB2 gene was positively correlated with the infiltration level of NK cells but was negatively correlated with the infiltration level of activated myeloid dendritic cells. The expression of the DCTN1 gene was positively correlated with the infiltration level of endothelial cells but was negatively correlated with

the infiltration level of common lymphoid progenitor cells (**Figure 5B** and **Table S7**). The above findings supported our conclusion that the hub genes played vital roles in the gastric tumorigenesis and indicated the potential relationships between the AS events and cell abundance.

Here, we postulated that the ESAS might be also involved in the immune cell infiltration. To test the hypothesis, the immune cell abundance in the EOGC was calculated using the ImmuCellAI database. Then, Spearman's rank correlation analyses were performed to indicate the relationship between the ESAS and immune cell infiltration in the EOGC. In total, 24 types of immune cells (including the infiltration score) were correlated with 77 ESAS events (Spearman's rank correlation coefficients > 0.4 or < -0.4, BH adjusted p-value < 0.05, **Figure 5C** and **Table S8**).



Network of the ESAS and Regulatory Factors

The AS events are regulated by various factors, including splicing factors (SFs) and APA core factors. However, it is unknown how the ESAS are regulated by these factors in the EOGC. To gain insights into this, we built the correlation network between the expression of 71 experimentally validated SFs, 22 APA core factors, and the PSI values of ESASs (27, 28).

In the network of the ESAS and SFs, 70 ESASs were associated with 25 SFs (Figure 6A and Table S9). All SFs were significantly correlated with at least five AS events. Moreover, one AS event could be regulated by up to 28 different SFs. We also constructed a network of APA core factors and TAPAs, which included seven APA core factors and ten ESAS events, respectively (Figure 6B and Table S9).

Clinically Relevant and Protein Modification Associated AS Events

There are few analyses to identify clinically relevant and protein modification associated AS events in the EOGC or GC. In this study, we identified 864 differentially expressed AS events from 577 genes between the EBV positive and negative tumor tissues (Figure 7A), 574 AS events from 533 genes associated with the MSI-High (MSI-H) and MSI-Low (MSI-L) tumor (Table S10).

The patients were subtyped according to the status of protein modifications, including phosphorylation and glycosylation, by a previous study (9). We found that 82 AS events from 60 genes were differentially expressed among the phosphorylation subtypes one to three, while the other 85 AS events from 65 genes were differentially expressed among the glycosylation subtypes one to three (Figures 7B, C and

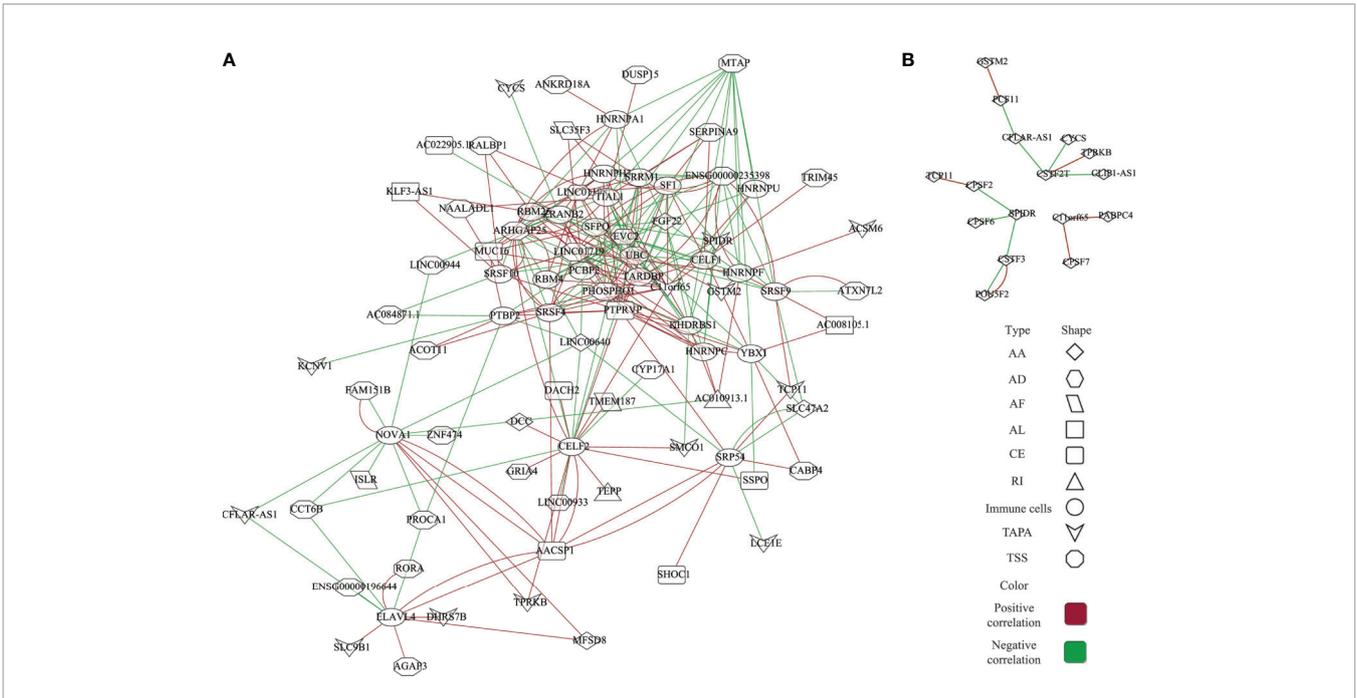


FIGURE 6 | Network of the ESAS and regulatory factors. **(A)** Spearman's rank correlation network of the ESAS and SFs; **(B)** Spearman's rank correlation network of the TAPA and APA core factors. Red: positive correlation; Green: Negative correlation.

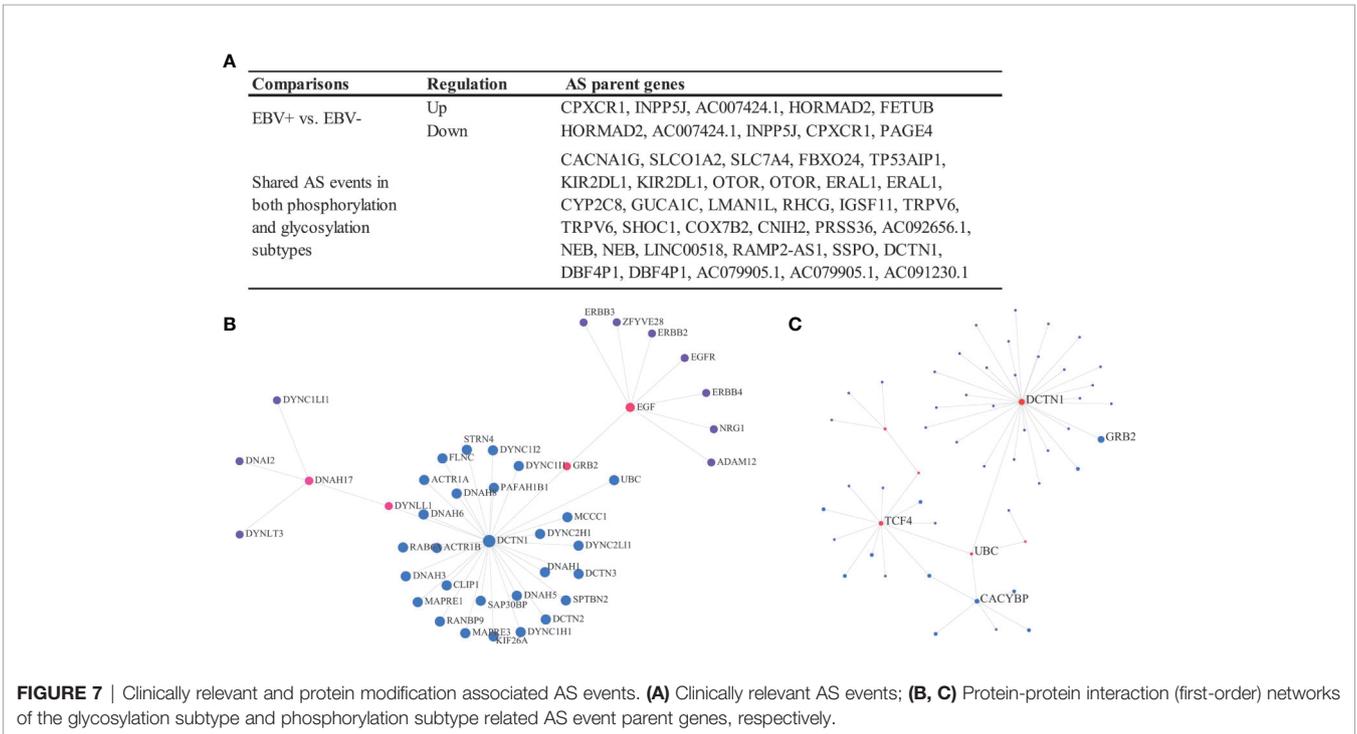


FIGURE 7 | Clinically relevant and protein modification associated AS events. **(A)** Clinically relevant AS events; **(B, C)** Protein-protein interaction (first-order) networks of the glycosylation subtype and phosphorylation subtype related AS event parent genes, respectively.

Table S10). In addition, a total of 27 AS events were dysregulated in both the phosphorylation and glycosylation subtypes (**Figure 7A**). The DCTN1 gene is a key hub gene in both phosphorylation and glycosylation networks.

DISCUSSION

Investigation of the AS events using the TCGA dataset has been a promising method and resulted in better understanding of the

roles of the mRNA processing in the malignancy diseases (33–35). However, it remains challenging to reveal the roles of the AS events in certain subtypes of cancer (7). The unique subtype of cancer can be contributed by the AS events during the tumorigenesis. For example, the *DOCK5* gene variant was identified as an oncogenic isoform in the HPV-negative head and neck squamous cell carcinoma (36). Herein, we conducted a systematic study to profile the AS events in the EOGC. In addition, we elucidated the roles of ESAS events with the splicing factors, APA core factors, immune cell infiltration, and protein modifications.

The distribution of the AS events in the gastric cancer was tissue-specific and related to the prognosis of patients (37). Furthermore, the expression patterns of the AS events in gastric cancer could be altered due to the Epstein-Barr virus infection (5). In this study, the RNA-Seq data with the protein modification status for subjects obtained from a published report were analyzed (9). In addition, we used Whippet to map the fastq format file to the reference genome, which generated two additional AS patterns (TSS and TAPA) than SpliceSeq (38) or MATS (39). These two approaches made the results more accurate and comprehensive than studies based on the TCGA SpliceSeq database (40).

According to our results, 267 ESAS, 6152 gene isoforms, and 4809 genes were aberrantly expressed in the EOGC patients. The parent genes of the ESAS were significantly enriched in the GO and KEGG. Further analyses indicated that the *UBC*, *NEK2*, *EPHB2*, and *DCTN1* genes were the hub genes in PPI networks. It was reported that the *UBC* gene regulated the cell ubiquitin under normal and stressful conditions and the TSS events were identified within the promoter region of the *UBC* gene (41). The *NEK2* protein was considered a splicing factor kinase through phosphorylation of the splicing factors, including the oncogenic *SRSF1* protein (42). Alternative splicing and alternative polyadenylation encoded a variant of the *EPHB2* gene which is a member of the EPH receptor protein-tyrosine kinase (43). The genetic structure variability of the *DCTN1* gene was involved in the development of the limb-girdle muscular dystrophy (44) and neuron differentiation (45). These reports might shed light on the roles of the AS events and hub genes in the gastric cancer pathogenesis.

The distribution of tumor-infiltrating lymphocytes in the gastric cancer was correlated with the tumor histological type and clinical outcome (46). On the basis of the immune infiltration data, the immunoscore, a prognostic signature, was established for prognostic predictions of gastric cancer (47, 48). Our results showed that the hub genes and ESAS were correlated with distinct immune cell populations in the EOGC. For example, the *TRIM45* gene encodes a protein to regulate the MAPK and NF- κ B pathways which may inhibit the cancer cell proliferation (49). In our study, the TSS events of the *TRIM 45* gene were positively correlated with macrophages but negatively correlated with NK T cells. Thus, we postulated that the alternative splicing of the genes might have potential roles in regulating the abundance of certain immune cells in the EOGC.

Two major types of regulatory factors, SFs and APA core factors, participated in the modulation of the AS events (15, 50). Hence, we

built two networks (SFs vs. ESAS and APA core factors vs. TAPA) based on the Spearman's rank correlation. We found that the dysregulation of SFs was highly associated with the ESAS expression and the APA core factors were linked to the TAPA events, which is consistent with the findings in previous reports (51, 52). Our results suggests that the SF and APA core factor contributes to the post-transcriptional mRNA processing of the EOGC.

Alternative splicing modulates the protein modification such as the protein phosphorylation (53) and protein glycosylation (54). The alternative splicing might modulate the protein modification through the isoform switch and change of the binding domain for N-linked glycosylation (55). One CD44 isoform has been shown to activate the RAS through the phosphorylation of the Erk (56). In addition, the small molecule, SM08502, reduced the phosphorylation of the splicing factors and generated the isoforms of the Wnt pathway genes which inhibited the gastrointestinal tumors (57). The reports about the alternative splicing and protein glycosylation are limited. Several studies focused on the CD44 variant 9 in the gastric cancer (58) and variant 6 in the colon cancer (59).

To our knowledge, the current study is the first to conduct a systematic analysis of the alternative splicing and protein modification based on the RNA sequence and mass spectrometry data in gastric cancer. The differential expressions of the AS events among the subtypes of the protein modification as well as EBV status and MSI status in the EOGC indicated that the AS events of the EOGC are subtype-specific. However, the shared AS events among different subtypes might demonstrate the close association of the subgroups defined by the protein modification or viral infection in the EOGC.

In conclusion, implementation of the rigorous criteria ensured the identification of the specific AS events related to the EOGC. In total, 267 ESAS events, 6152 gene isoforms, and 4809 genes were identified and might play vital roles in the EOGC tumorigenesis. The hub genes of the PPI networks and ESAS events might be valuable in deciphering the immune microenvironment in the early-onset gastric carcinogenesis. In addition, the SF and APA core factor correlation networks revealed the underlying pathways of the splicing modulation. Furthermore, a comprehensive analysis of the subtype-specific AS events suggested certain connection between the protein modification and viral infection in the EOGC. The findings in this study might be valuable in the clinical diagnosis and prediction of the early-onset gastric cancer.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ebi.ac.uk/ena/browser/view/PRJNA508414>.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local

legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Designed the study and analyzed the data: JZ and LZ. Critical revision and structure suggestions: AG. All authors contributed to the article and approved the submitted version.

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
- Sung H, Siegel RL, Rosenberg PS, Jemal A. Emerging Cancer Trends Among Young Adults in the USA: Analysis of a Population-Based Cancer Registry. *Lancet Public Health* (2019) 4(3):e137–e47. doi: 10.1016/S2468-2667(18)30267-6
- Takatsu Y, Hiki N, Nunobe S, Ohashi M, Honda M, Yamaguchi T, et al. Clinicopathological Features of Gastric Cancer in Young Patients. *Gastric Cancer* (2016) 19(2):472–8. doi: 10.1007/s10120-015-0484-1
- Shi Y, Chen Z, Gao J, Wu S, Gao H, Feng G. Transcriptome-Wide Analysis of Alternative mRNA Splicing Signature in the Diagnosis and Prognosis of Stomach Adenocarcinoma. *Oncol Rep* (2018) 40(4):2014–22. doi: 10.3892/or.2018.6623
- Armero VES, Tremblay MP, Allaire A, Boudreault S, Martenon-Brodeur C, Duval C, et al. Transcriptome-Wide Analysis of Alternative RNA Splicing Events in Epstein-Barr Virus-Associated Gastric Carcinomas. *PLoS One* (2017) 12(5):e0176880. doi: 10.1371/journal.pone.0176880
- Ray D, Yun YC, Idris M, Cheng S, Boot A, Iain TBH, et al. A Tumor-Associated Splice-Isoform of MAP2K7 Drives Dedifferentiation in MBNL1-low Cancers Via JNK Activation. *Proc Natl Acad Sci* (2020) 117(28):16391–400. doi: 10.1073/pnas.2002499117
- Huang KK, Huang J, Wu JKL, Lee M, Tay ST, Kumar V, et al. Long-Read Transcriptome Sequencing Reveals Abundant Promoter Diversity in Distinct Molecular Subtypes of Gastric Cancer. *Genome Biol* (2021) 22(1):44. doi: 10.1186/s13059-021-02261-x
- Cho SY, Park JW, Liu Y, Park YS, Kim JH, Yang H, et al. Sporadic Early-Onset Diffuse Gastric Cancers Have High Frequency of Somatic Cdh1 Alterations, But Low Frequency of Somatic RhoA Mutations Compared With Late-Onset Cancers. *Gastroenterology* (2017) 153(2):536–49.e26. doi: 10.1053/j.gastro.2017.05.012
- Mun DG, Bhin J, Kim S, Kim H, Jung JH, Jung Y, et al. Proteogenomic Characterization of Human Early-Onset Gastric Cancer. *Cancer Cell* (2019) 35(1):111–24.e10. doi: 10.1016/j.ccell.2018.12.003
- Modrek B, Lee C. A Genomic View of Alternative Splicing. *Nat Genet* (2002) 30(1):13–9. doi: 10.1038/ng0102-13
- Black DL. Protein Diversity From Alternative Splicing: A Challenge for Bioinformatics and Post-Genome Biology. *Cell* (2000) 103(3):367–70. doi: 10.1016/S0092-8674(00)00128-8
- Li X, Zhang C, Gong T, Ni X, Li J, Zhan D, et al. A Time-Resolved Multi-Omic Atlas of the Developing Mouse Stomach. *Nat Commun* (2018) 9(1):4910. doi: 10.1038/s41467-018-07463-9
- Milne AN, Carvalho R, Morsink FM, Musler AR, de Leng WW, Ristimaki A, et al. Early-Onset Gastric Cancers Have a Different Molecular Expression Profile Than Conventional Gastric Cancers. *Mod Pathol* (2006) 19(4):564–72. doi: 10.1038/modpathol.3800563
- Jia X, Yuan S, Wang Y, Fu Y, Ge Y, Ge Y, et al. The Role of Alternative Polyadenylation in the Antiviral Innate Immune Response. *Nat Commun* (2017) 8(1):14605. doi: 10.1038/ncomms14605

FUNDING

The work was partially supported by the National Cancer Institute of the National Institutes of Health (R15CA213103) and Texas A&M University T3 grant (247099).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.640272/full#supplementary-material>

- Xiang Y, Ye Y, Lou Y, Yang Y, Cai C, Zhang Z, et al. Comprehensive Characterization of Alternative Polyadenylation in Human Cancer. *J Natl Cancer Inst* (2018) 110(4):379–89. doi: 10.1093/jnci/djx223
- Fischl H, Neve J, Wang Z, Patel R, Louey A, Tian B, et al. hnRNPC Regulates Cancer-Specific Alternative Cleavage and Polyadenylation Profiles. *Nucleic Acids Res* (2019) 47(14):7580–91. doi: 10.1093/nar/gkz461
- Lai DP, Tan S, Kang YN, Wu J, Ooi HS, Chen J, et al. Genome-Wide Profiling of Polyadenylation Sites Reveals a Link Between Selective Polyadenylation and Cancer Metastasis. *Hum Mol Genet* (2015) 24(12):3410–7. doi: 10.1093/hmg/ddv089
- Elkon R, Ugalde AP, Agami R. Alternative Cleavage and Polyadenylation: Extent, Regulation and Function. *Nat Rev Genet* (2013) 14(7):496–506. doi: 10.1038/nrg3482
- Sterne-Weiler T, Weatheritt RJ, Best AJ, Ha KCH, Blencowe BJ. Efficient and Accurate Quantitative Profiling of Alternative Splicing Patterns of Any Complexity on a Laptop. *Mol Cell* (2018) 72(1):187–200.e6. doi: 10.1016/j.molcel.2018.08.018
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
- McCarthy DJ, Chen Y, Smyth GK. Differential Expression Analysis of Multifactor RNA-Seq Experiments With Respect to Biological Variation. *Nucleic Acids Res* (2012) 40(10):4288–97. doi: 10.1093/nar/gks042
- Conway JR, Lex A, Gehlenborg N. UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties. *Bioinformatics* (2017) 33(18):2938–40. doi: 10.1093/bioinformatics/btx364
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
- Xia J, Gill EE, Hancock RE. NetworkAnalyst for Statistical, Visual and Network-Based Meta-Analysis of Gene Expression Data. *Nat Protoc* (2015) 10(6):823–44. doi: 10.1038/nprot.2015.052
- Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res* (2017) 77(21):e108–e10. doi: 10.1158/0008-5472.CAN-17-0307
- Miao YR, Zhang Q, Lei Q, Luo M, Xie GY, Wang H, et al. Immucellai: A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and Its Application in Cancer Immunotherapy. *Adv Sci (Weinh)* (2020) 7(7):1902880. doi: 10.1002/advs.201902880
- Giulietti M, Piva F, D'Antonio M, D'Onorio De Meo P, Paoletti D, Castrignano T, et al. SpliceAid-F: A Database of Human Splicing Factors and Their RNA-binding Sites. *Nucleic Acids Res* (2013) 41(Database issue):D125–31. doi: 10.1093/nar/gks997
- Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR3rd, et al. Molecular Architecture of the Human Pre-mRNA 3' Processing Complex. *Mol Cell* (2009) 33(3):365–76. doi: 10.1016/j.molcel.2008.12.028
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303
- Team RC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria: Title: R Core Team (2020).

31. Perle M, Shumate A, Perle G, Varabyou A, Breitwieser FP, Chang YC, et al. CHES: A New Human Gene Catalog Curated From Thousands of Large-Scale RNA Sequencing Experiments Reveals Extensive Transcriptional Noise. *Genome Biol* (2018) 19(1):208. doi: 10.1186/s13059-018-1590-2
32. Law CW, Chen Y, Shi W, Smyth GK. Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-seq Read Counts. *Genome Biol* (2014) 15(2):R29. doi: 10.1186/gb-2014-15-2-r29
33. Zhang Y, Yan L, Zeng J, Zhou H, Liu H, Yu G, et al. Pan-Cancer Analysis of Clinical Relevance of Alternative Splicing Events in 31 Human Cancers. *Oncogene* (2019) 38(40):6678–95. doi: 10.1038/s41388-019-0910-7
34. Li Y, Sun N, Lu Z, Sun S, Huang J, Chen Z, et al. Prognostic Alternative mRNA Splicing Signature in Non-Small Cell Lung Cancer. *Cancer Lett* (2017) 393:40–51. doi: 10.1016/j.canlet.2017.02.016
35. Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, et al. Dynamic Analyses of Alternative Polyadenylation From RNA-seq Reveal a 3'-UTR Landscape Across Seven Tumour Types. *Nat Commun* (2014) 5(1):5274. doi: 10.1038/ncomms6274
36. Liu C, Guo T, Xu G, Sakai A, Ren S, Fukusumi T, et al. Characterization of Alternative Splicing Events in HPV-Negative Head and Neck Squamous Cell Carcinoma Identifies an Oncogenic Dock5 Variant. *Clin Cancer Res* (2018) 24(20):5123–32. doi: 10.1158/1078-0432.CCR-18-0752
37. Lin P, He RQ, Ma FC, Liang L, He Y, Yang H, et al. Systematic Analysis of Survival-Associated Alternative Splicing Signatures in Gastrointestinal Pan-Adenocarcinomas. *EBioMedicine* (2018) 34:46–60. doi: 10.1016/j.ebiom.2018.07.040
38. Ryan MC, Cleland J, Kim R, Wong WC, Weinstein JN. SpliceSeq: A Resource for Analysis and Visualization of RNA-Seq Data on Alternative Splicing and its Functional Impacts. *Bioinformatics* (2012) 28(18):2385–7. doi: 10.1093/bioinformatics/bts452
39. Shen S, Park JW, Huang J, Dittmar KA, Lu ZX, Zhou Q, et al. MATS: A Bayesian Framework for Flexible Detection of Differential Alternative Splicing From RNA-Seq Data. *Nucleic Acids Res* (2012) 40(8):e61. doi: 10.1093/nar/gkr1291
40. Ryan M, Wong WC, Brown R, Akbani R, Su X, Broom B, et al. TcspliceSeq a Compendium of Alternative mRNA Splicing in Cancer. *Nucleic Acids Res* (2016) 44(D1):D1018–22. doi: 10.1093/nar/gkv1288
41. Bianchi M, Giacomini E, Crinelli R, Radici L, Carloni E, Magnani M. Dynamic Transcription of Ubiquitin Genes Under Basal and Stressful Conditions and New Insights Into the Multiple UBC Transcript Variants. *Gene* (2015) 573(1):100–9. doi: 10.1016/j.gene.2015.07.030
42. Naro C, Barbagallo F, Chieffi P, Bourgeois CF, Paronetto MP, Sette C. The Centrosomal Kinase NEK2 is a Novel Splicing Factor Kinase Involved in Cell Survival. *Nucleic Acids Res* (2014) 42(5):3218–27. doi: 10.1093/nar/gkt1307
43. Tang XX, Pleasure DE, Brodeur GM, Ikegaki N. A Variant Transcript Encoding an Isoform of the Human Protein Tyrosine Kinase EPHB2 is Generated by Alternative Splicing and Alternative Use of Polyadenylation Signals. *Oncogene* (1998) 17(4):521–6. doi: 10.1038/sj.onc.1201960
44. Tokito MK, Holzbaur EL. The Genomic Structure of DCTN1, A Candidate Gene for Limb-Girdle Muscular Dystrophy (LGMD2B). *Biochim Biophys Acta* (1998) 1442(2-3):432–6. doi: 10.1016/S0167-4781(98)00195-X
45. Song Y, Botvinnik OB, Lovci MT, Kakaradov B, Liu P, Xu JL, et al. Single-Cell Alternative Splicing Analysis With Expedition Reveals Splicing Dynamics During Neuron Differentiation. *Mol Cell* (2017) 67(1):148–61.e5. doi: 10.1016/j.molcel.2017.06.003
46. Pernot S, Terme M, Radosevic-Robin N, Castan F, Badoual C, Marcheteau E, et al. Infiltrating and Peripheral Immune Cell Analysis in Advanced Gastric Cancer According to the Lauren Classification and Its Prognostic Significance. *Gastric Cancer* (2020) 23(1):73–81. doi: 10.1007/s10120-019-00983-3
47. Zeng D, Zhou R, Yu Y, Luo Y, Zhang J, Sun H, et al. Gene Expression Profiles for a Prognostic Immunoscore in Gastric Cancer. *Br J Surg* (2018) 105(10):1338–48. doi: 10.1002/bjs.10871
48. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. *Ann Surg* (2018) 267(3):504–13. doi: 10.1097/SLA.0000000000002116
49. Shibata M, Sato T, Nukiwa R, Ariga T, Hatakeyama S. TRIM45 Negatively Regulates NF-kappaB-mediated Transcription and Suppresses Cell Proliferation. *Biochem Biophys Res Commun* (2012) 423(1):104–9. doi: 10.1016/j.bbrc.2012.05.090
50. Watermann DO, Tang Y, Zur Hausen A, Jager M, Stamm S, Stickeler E. Splicing Factor Tra2-beta1 is Specifically Induced in Breast Cancer and Regulates Alternative Splicing of the CD44 Gene. *Cancer Res* (2006) 66(9):4774–80. doi: 10.1158/0008-5472.CAN-04-3294
51. Wang R, Zheng D, Yehia G, Tian B. A Compendium of Conserved Cleavage and Polyadenylation Events in Mammalian Genes. *Genome Res* (2018) 28(10):1427–41. doi: 10.1101/gr.237826.118
52. Listerman I, Sapra AK, Neugebauer KM. Cotranscriptional Coupling of Splicing Factor Recruitment and Precursor Messenger RNA Splicing in Mammalian Cells. *Nat Struct Mol Biol* (2006) 13(9):815–22. doi: 10.1038/nsmb1135
53. Venables JP. Aberrant and Alternative Splicing in Cancer. *Cancer Res* (2004) 64(21):7647–54. doi: 10.1158/0008-5472.CAN-04-1910
54. Bach MA, Roberts CT Jr, Smith EP, LeRoith D. Alternative Splicing Produces Messenger RNAs Encoding Insulin-Like Growth Factor-I Prohormones That are Differentially Glycosylated In Vitro. *Mol Endocrinol* (1990) 4(6):899–904. doi: 10.1210/mend-4-6-899
55. Reale MA, Hu G, Zafar AI, Getzenberg RH, Levine SM, Fearon ER. Expression and Alternative Splicing of the Deleted in Colorectal Cancer (EmDCC</math> Gene in Normal and Malignant Tissues. *Cancer Res* (1994) 54(16):4493.
56. Cheng C, Yaffe MB, Sharp PA. A Positive Feedback Loop Couples Ras Activation and CD44 Alternative Splicing. *Genes Dev* (2006) 20(13):1715–20. doi: 10.1101/gad.1430906
57. Tam BY, Chiu K, Chung H, Bossard C, Nguyen JD, Creger E, et al. The CLK Inhibitor SM08502 Induces Anti-Tumor Activity and Reduces Wnt Pathway Gene Expression in Gastrointestinal Cancer Models. *Cancer Lett* (2020) 473:186–97. doi: 10.1016/j.canlet.2019.09.009
58. Moreira IB, Pinto F, Gomes C, Campos D, Reis CA. Impact of Truncated O-Glycans in Gastric-Cancer-Associated Cd44v9 Detection. *Cells* (2020) 9(2):264. doi: 10.3390/cells9020264
59. Indinnimeo M, Cicchini C, Giarnieri E, Stazi A, Mingazzini PL, Stipa V. Evaluation of CD44 Variant 6 Expression and Clinicopathological Factors in Pulmonary Metastases From Colon Carcinoma. *Oncol Rep* (2003) 10(6):1875–7. doi: 10.3892/or.10.6.1875

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.

Copyright © 2021 Zhang, Goel 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.