Computational and Structural Biotechnology Journal 21 (2023) 3841–3853

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Comparative transcriptome characterization of esophageal squamous cell carcinoma and adenocarcinoma

Xianfeng Li [a,b,c,d,1], Yan Wang [a,1], Qingjie Min [a,1], Weimin Zhang [a], Huajing Teng [e], Chao Li [f], Kun Zhang [f], Leisheng Shi [g], Bin Wang [b,c,d], Qimin Zhan [a,*]

[a] Key laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China
[b] Department of Gastroenterology and Chongqing Key Laboratory of Digestive Malignancies, Daping Hospital, Army Medical University (Third Military Medical University), 10# Changjiang Branch Road, Yuzhong District, Chongqing 400042, People's Republic of China
[c] Institute of Pathology and Southwest Cancer Center, Key Laboratory of Tumor Immunopathology of Ministry of Education of China, Southwest Hospital, Army Medical University (Third Military Medical University), Chongqing 400038, People's Republic of China
[d] Jinfeng Laboratory, Chongqing 401329, People's Republic of China
[e] Key laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China
[f] Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou 325035, China
[g] Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:
Esophageal squamous cell carcinoma
Esophageal adenocarcinoma
Transcriptome

ABSTRACT

Background: Esophageal cancers are primarily categorized as esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC). While various (epi) genomic alterations associated with tumor development in ESCC and EAC have been documented, a comprehensive comparison of the transcriptomes in these two cancer subtypes remains lacking.

Methods: We collected 551 gene expression profiles from publicly available sources, including normal, ESCC, and EAC tissues or cell lines. Subsequently, we conducted a systematic analysis to compare the transcriptomes of these samples at various levels, including gene expression, promoter activity, alternative splicing (AS), alternative polyadenylation (APA), and gene fusion.

Results: Seven distinct cluster gene expression patterns were identified among the differentially expressed genes in normal, ESCC, and EAC tissues. These patterns were enriched in the PI3K-Akt signaling pathway and the activation of extracellular matrix organization and exhibited repression of epidermal development. Notably, we observed additional genes or unique expression levels enriched in these shared pathways and biological processes related to tumor development and immune activation. In addition to the differentially expressed genes, there was an enrichment of lncRNA co-expression networks and downregulation of promoter activity associated with the repression of epidermal development in both ESCC and EAC. This indicates a common feature between these two cancer subtypes. Furthermore, differential AS and APA patterns in ESCC and EAC appear to partially affect the expression of host genes associated with bacterial or viral infections in these subtypes. No gene fusions were observed between ESCC and EAC, thus highlighting the distinct molecular mechanisms underlying these two cancer subtypes.

Conclusions: We conducted a comprehensive comparison of ESCC and EAC transcriptomes and uncovered shared and distinct transcriptomic signatures at multiple levels. These findings suggest that ESCC and EAC may exhibit common and unique mechanisms involved in tumorigenesis.

Abbreviations: **ESCC**, Esophageal squamous cell carcinoma; **EAC**, Esophageal adenocarcinoma; **LncRNA**, Long non-coding RNA; **APA**, Alternative polyadenylation; **AS**, Alternative splicing; **TCGA**, The Cancer Genome Atlas; **GTEx**, The Genotype-Tissue Expression project; **CCLE**, Broad Institute Cancer Cell Line Encyclopedia; **NCBI SRA**, The Sequence Read Archive of National Center for Biotechnology Information; **GENCODE**, The reference human genome annotation for The ENCODE Project; **FDR**, False discovery rate; **TPM**, Transcripts per million; **FFPM**, Fusion fragments per million; **RBP**, RNA-binding protein; **ECM**, Extracellular matrix; **TME**, tumorigenic microenvironment; **UTR,**, untranslated region.

* Corresponding author.
E-mail address: zhanqimin@bjmu.edu.cn (Q. Zhan).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2023.07.030
Received 9 December 2022; Received in revised form 20 July 2023; Accepted 22 July 2023
Available online 25 July 2023

1. Introduction

Esophageal cancers ranked seventh cancer incidence (604,000 new cases) and sixth in cancer mortality (544,000 deaths) in 2020 with a less than 20% 5-year survival rate for all patients globally [1,2]. Approximately 70% of esophageal cancer cases occur in men at 2–3-fold higher incidence and mortality rates than in women. Esophageal cancers are typically classified as either esophageal squamous cell carcinoma (ESCC) or esophageal adenocarcinoma (EAC) based on their histology [1,3–5]. ESCC is the most common type and accounts for almost 90% of esophageal cancer cases, and typically occurs in the upper- and mid-esophagus. It is primarily observed in high-risk areas in middle- and low-income regions, such as Asia, Southern Africa, Eastern Africa, Northern Europe, and South Central Asia. Suspected risk factors in these areas include heavy drinking, smoking, thermal injury, and dietary components, such as nutritional deficiencies and nitrosamines [6–9]. While EACs tend to be located in the lower esophagus near the gastric junction (GEJ) and possess a glandular structure [10], incidences of these cancers are rising rapidly in high-income countries (such as the United States, Australia, France, and the United Kingdom), in part due to increasing risk factors, such as obesity and gastric reflux, that result in the precursor state known as Barrett's esophagus (BE) [2]. Repeated insults to the esophageal mucosa by these risk factors result in precancerous lesions with low- and high-grade squamous dysplasia or metaplastic epithelial dysplasia that lead to invasive ESCC or EAC [1]. Additionally, fungal infections in the oral cavity, including the periodontal pathogens, *Tannerella forsythia* and *Porphyromonas gingivalis*, are associated with a higher risk of ESCC and EAC, respectively [11,12].

Esophageal cancer is often diagnosed at an advanced stage, thus leading to limited treatment options, high rates of recurrence and metastasis, and poor survival outcomes. However, recent clinical trials have demonstrated that neoadjuvant chemoradiotherapy (CRT) or perioperative chemotherapy, can help reduce tumor size, improve local control, and increase the likelihood of achieving R0 resection, ultimately resulting in significant improvements in overall and disease-free survival rates [13,14]. Furthermore, the use of immune checkpoint inhibitors (ICIs) as adjuvant therapy after neoadjuvant CRT and surgery has yielded promising results in regard to improving disease-free survival in patients who have undergone resection for esophageal or gastroesophageal junction cancer [15]. However, these clinical studies revealed that EAC and ESCC exhibit different responses to various neoadjuvant therapies based on their histologic type [13–16]. Further investigation is necessary to elucidate the tumor microenvironment (TME) and underlying molecular mechanisms of the two subtypes of esophageal cancer to improve clinical outcomes. As ESCC and EAC exhibit different histology and epidemiology, numerous studies have been conducted to identify their distinct molecular signatures, including DNA aberrations, epigenetic alterations, and gene expression patterns [9,17–24]. Recently, integrated genomic characterizations, including somatic copy number alteration (SCNA), DNA methylation, mRNA, and microRNA expression, revealed that ESCC was more closely related to head and neck squamous carcinoma (HNSCC) and lung squamous cell carcinoma (SQCC) and harbored more frequent genomic amplifications of *CCND1*, *SOX2/TP63*, and *KDM6A* deletions than did EAC [4,8,10]. EAC is highly similar to the chromosomally unstable variant of gastric adenocarcinoma with increased amplification of *ERBB2, VEGFA*, *GATA4*, and *GATA6* [10].

Currently, RNA-Seq techniques are used to generate gene expression profiles to identify abnormal expression patterns in ESCC and EAC [8,10,25,26]. For example, lncRNA625 was identified as a novel regulator of ESCC cell proliferation, invasion, and migration [25]. The mRNA levels of *CTSL, COL17A1, KLF4*, and *E2F3* in EAC cells differed from those in BE [26]. Nevertheless, a thorough comparison of the transcriptomic signatures of these two cancer subtypes remains ambiguous.

In this study, we collected and analyzed 551 transcriptomes from both normal and cancerous esophageal tissues, including 316 normal tissues , 104 EAC and 131 ESCC from the Genotype-Tissue Expression (GTEx) [27], the Sequence Read Archive of National Center for Biotechnology Information (NCBI SRA) [28], the Broad Institute Cancer Cell Line Encyclopedia (CCLE) [29], and the ESCA cohort of The Cancer Genome Atlas (TCGA) projects [30]. Subsequently, we conducted a comprehensive comparison between the two epidemiologically and biologically distinct cancers at coding/noncoding RNA expression, promoter activity, AS, APA, and gene fusion levels. Our findings revealed shared and unique transcriptomic signatures in regard to the tumorigenesis of these two cancer subtypes.

2. Materials and methods

2.1. Data collection and pre-processing

We collected and analyzed 551 gene expression profiles of esophageal tissue using RNA-Seq from the GTEx, NCBI SRA, CCLE, and ESCA cohorts of the TCGA project. Among the 551 transcriptomes, per-gene read counts for 465 of them were obtained from the UCSC Xena platform, including 271 normal esophageal tissues from GTEx 194 tissues from TCGA (13 normal esophageal tissue, 89 EAC, and 92 ESCC) [31]. The other accessible raw sequencing datasets consisting of 24 ESCC and 3 EAC cell lines in CCLE, 15 paired ESCC with matched adjacent normal samples, and 12 EAC and 17 normal tissues with accession IDs SRP064894 [25] and ERP013206 [26], respectively, were downloaded. Low quality reads and adaptors were removed from the raw sequencing datasets using fastp (v0.23.3) [32]. Clean reads were aligned to the hg38 genome using HISAT2 (v2.2.1) [33]. FeatureCounts (v2.0.3) [34] was used to calculate the read counts of mRNA and lncRNAs in GENCODE (v23) [35]. Furthermore, all 551 read count matrix were merged, and ComBat-seq [36] was utilized to eliminate batch effects arising from data sourced from different studies.

2.2. Gene expression profiles

Genes with counts per million of < 1 in more than 1/3 of samples were excluded. Differential expression between cancer and normal samples was determined using edgeR (v3.40.2) [37] with a cutoff of false discovery rate (FDR) of < 0.05 and an absolute fold change of > 2. An in-house script was used to convert the read counts to transcripts per million (TPM). Furthermore, ImmuCellAI (http://bioinfo.life.hust.edu.cn/ImmuCellAI) [38] was used to predict the abundance of 24 different types of immune cells based on a gene set of single samples. To cluster the differentially expressed genes, ClusterGVis (v0.1.0) (https://github.com/junjunlab/ClusterGVis) was employed. Co-expressed lncRNA-mRNA pairs were identified based on Pearson correlation, with a threshold of $r > 0.75$ and FDR < 0.05. The co-expression network was visualized using Cytoscape (v3.9.1) [39]. To determine the prognostic effects of gene expression, the R packages survminer (v0.4.9) [40] and survival (v3.5.5) [41] were adopted to calculate the log-rank p-value using the Kaplan–Meier method and hazard ratios using the Cox proportional hazards model based on gene expression levels and the clinical information of EAC or ESCC from the ESCA cohort of the TCGA project.

2.3. Promoter activity analysis and alternative promoter detection

To estimate the promoter activity, proActiv (v1.1.18) [42] was used to calculate the read count of each promoter from the raw ESCC (SRP064894) and EAC (ERP013206) sequencing datasets. Subsequently, the read count matrix of the promoters was imported into edgeR, and the differential activity of promoters between normal and tumor groups was defined with a cutoff of FDR < 0.05 and an absolute fold change > 2. A gene featuring two or more promoters with differential activity and differential isoform expression is considered a candidate for alternative promoters.

2.4. Alternative splicing analysis

To identify alternative splicing, rMATS (v4.0.2) [43] was adopted for ESCC- and EAC-aligned bam files. Alternative splicing was determined based on splicing patterns between normal and tumor groups with a cutoff of FDR < 0.05 and IncLevelDifference > 0.1. Subsequently, alternative splicing events from multiple samples were assessed using ggsashimi (v1.1.5) [44].

2.5. Alternative polyadenylation detection

The abundance of typical polyadenylation signal at the 3′UTR was calculated using QAPA (v1.3.3) [45]. The differential usage of polyadenylation was detected using edgeR with a cutoff of FDR < 0.05, and an absolute fold change > 2. Genes with differential polyadenylation usage and two or more polyadenylation sites (PAS) were selected for APA detection. Shortened or lengthened 3′UTRs were defined as greater than 2-fold tumor/normal ratio of proximal PAS (pPAS) and distal PAS (dPAS).

2.6. Gene fusion analysis

Gene fusions were identified from the RNA-Seq datasets using STAR-Fusion (v1.12.0) [46]. "FusionFilter" was used to filter several steps, including paralog fusion, promiscuous fusion, "Red herring," and fusion expression FFPM < 0.1. Further, gene fusions were annotated using

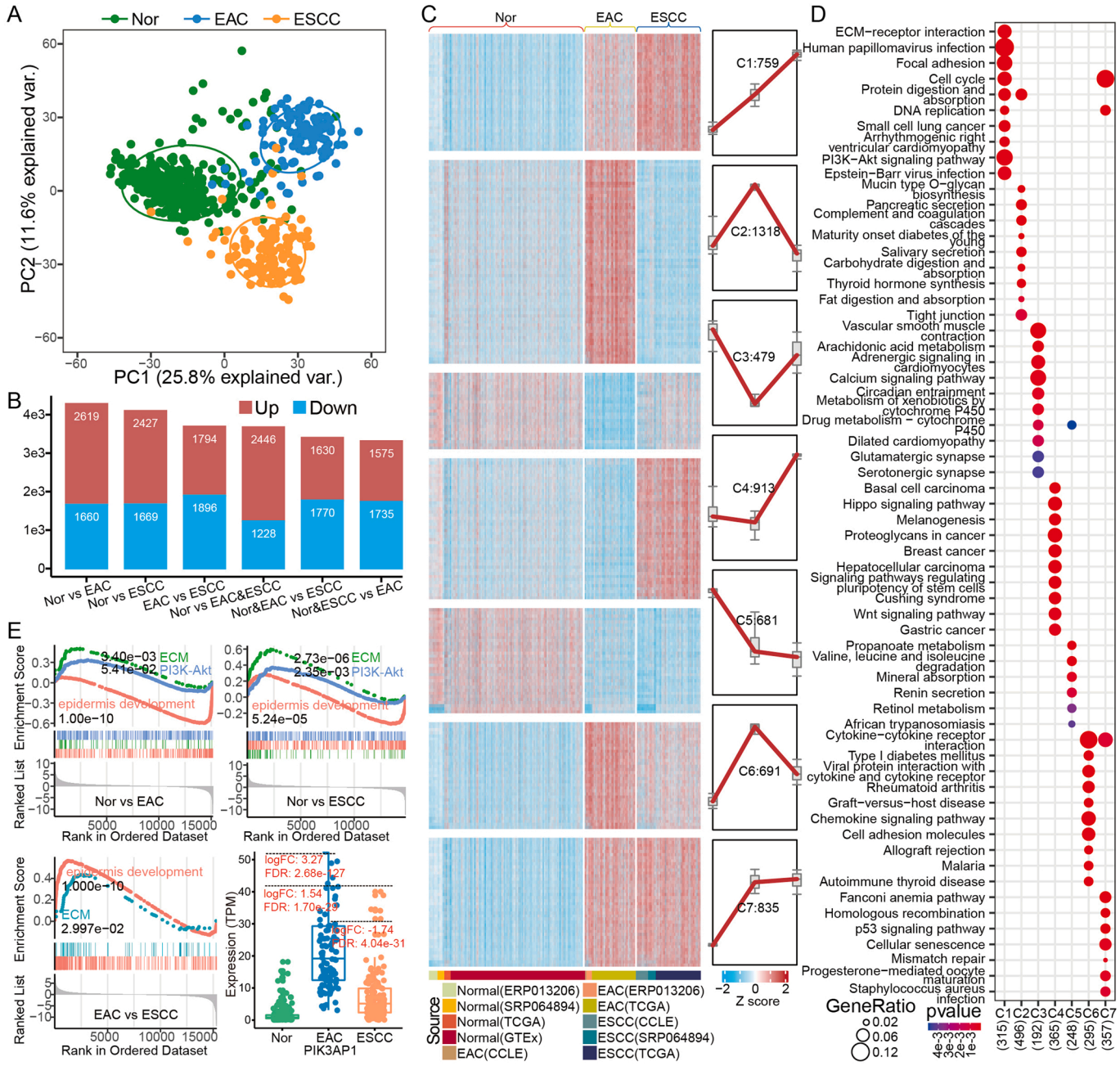

Fig. 1. : Differential mRNA expression patterns among normal tissue, EAC and ESCC. **A).** PCA of the transcriptome of normal tissue, EAC, and ESCC. **B)** Differentially expressed genes between different subgroups. **C).** Seven clusters based on expression patterns among the three groups. **D).** KEGG pathway enrichment of different group genes based on seven clusters in C. **E).** GSEA enriched analysis of ECM, PI3K-Akt, and epidermis development. *PIK3AP1* expression levels among normal tissue, EAC and ESCC are indicated in the right corner.

"FusionAnnotator" with 26 public gene fusion databases integrated into this package. Both "FusionFilter" and "FusionAnnotator" were packaged in the STAR-Fusion pipeline. Finally, gene fusions were visualized using chimeraviz (v1.26.0) [47] and Integrative Genomics Viewer (IGV v2.12.3) [48].

2.7. Motifs and function enrichment analysis

The script"findMotifs.pl" from homer [49] was implemented to perform motif enrichment analysis for transcription factors. RNA-binding protein (RBP) motif enrichment around alternative splicing sites was performed using the web server rMAPS (http://rmaps.cecsresearch.org) [50]. For RNA motif enrichment, DREME [51] (a MEME suite, v5.5.3, https://meme-suite.org) was used to identify RNA motifs, and TOMTOM [52] was used to quantify the similarity between the identified RNA motif and known RBP or miRNA target motifs. CentriMo [53] was used to visualize the signals of the predicted motifs around PAS site. Gene function enrichment analysis was performed using clusterProfiler (v4.6.2) [54].

3. Results

3.1. Abnormally expressed genes displayed both similar and specific functions in EAC and ESCC

We analyzed 551 gene expression profiles of normal and cancerous esophageal tissues obtained using RNA-Seq from public databases, including 316 normal esophageal, 104 EAC, and 131 ESCC samples. After batch-effect removal, principal component analysis (PCA) indicated that the three groups could be separated from each other (Fig. 1A). Subsequently, we identified 4309, 4092, 3599, 3559, 3211, and 3313 differentially expressed mRNAs in the normal tissue (Nor) vs. EAC, Nor vs. ESCC, EAC vs. ESCC, Nor vs. EAC&ESCC, Nor and EAC vs. ESCC, and Nor and ESCC vs. EAC groups, respectively (Fig. 1B). Among the total 6854 differentially expressed genes, seven clusters exhibiting distinct patterns among the three groups were identified (Fig. 1C). The differentially expressed genes in ESCC and EAC possessed both shared and distinct functions, as revealed by biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses (Fig. 1D; Figure S1A). Genes in cluster C1 were enriched in several tumor-related pathways, including focal adhesion, cell cycle, extracellular matrix (ECM), and the PI3K-Akt signaling pathway. Consistent with genomic alterations [10], cell growth and proliferation stimulated by the PI3K-Akt signaling pathway in ESCC and EAC may be significantly activated at high expression levels (Fig. 1 E). The expression of *PIK3AP1* was significantly higher in EAC than in normal esophageal samples and ESCC (Fig. 1E). ECM-receptor interactions (Fig. 1D, E) and ECM-related biological processes were enriched in both ESCC and EAC, thus suggesting the TME may be affected by the high ECM gene expression [55]. Furthermore, DNA repair-related pathways, such as the p53 signaling pathway and mismatch repair, were enriched in cluster C7, with high expression levels in both EAC and ESCC. Notably, genes in C7 and C6 were enriched in cytokine–cytokine receptor interactions, whereas the C6 genes were enriched in immune-related pathways, such as chemokine signaling pathway or T cell activation biology processes (Fig. 1D; Figure S1A). This suggests that the differential tumor microenvironments of the two cancer subtypes may result in varying responses to immunotherapy.

Compared with that in normal esophageal tissue, the predicted immune cell abundance in EAC and ESCC was significantly higher and lower respectively (Figure S1B, C). For example, the abundance of type 1 regulatory T cells (Tr1), dendritic cells (DC), and macrophages was higher in both EAC and ESCC, and the abundance of the three cell types in ESCC was the highest among the three groups (Figure S1C). This indicated that EAC and ESCC may possess differential TMEs, even though both ECM-related genes were upregulated. Additionally, activation of both the PI3K-Akt signaling pathway and ECM-related pathways was enriched in ESCC and EAC (Fig. 1 D, E). Epidermal development was repressed in these two cancer subtypes (Fig. 1D, E; Figure S1A). ESCC and EAC exhibit distinct biological processes and pathways. For example, the Hippo and WNT signaling pathways were specifically enriched in cluster C4, with higher gene expression observed in ESCC (Fig. 1D, a; Figure S1A). Cluster C2 genes with higher expression levels in the EAC were enriched in digestion-related pathways and biological processes (Fig. 1D; Fig S1A).

We also identified 1159 differentially expressed lncRNAs in normal, EAC, and ESCC tissues (Fig. 2A). Among these lncRNAs, 35 genes that were involved in esophageal cancer were reported in the lncRNADisease database (http://www.rnanut.net/lncrnadisease) [56] (Fig. 2B). Furthermore, the co-expression networks of differentially expressed lncRNAs and mRNAs included 2923 (323 lncRNAs, 909 mRNAs) and 4980 (383 lncRNAs and 1218 mRNAs, Pearson's $r > 0.75$, FDR<0.05) in EAC and ESCC, respectively (Fig. 2C, D). Notably, *TINCR* and *HOTTIP* were ranked as the most connected in the EAC co-expressed network (Fig. 2C). The reduced expression of *TINCR* in the EAC may result in the loss of stability regulation of genes involved in the differentiation of the epidermal tissue [57]. *HOTTIP*, termed as 'HOXA transcript at the distal tip', was associated with various malignancies, including hepatocellular carcinoma, pancreatic cancer, gastric cancer and colorectal cancer [58]. Increased expression of *HOTTIP* may promotes cell proliferation in EAC. Meanwhile, several other lncRNAs, such as the antisense gene *TMPO-AS1*, ranked at the top (Fig. 2D), and they promote esophageal squamous cell carcinoma progression by forming biomolecular condensates with FUS and p300 to regulate *TMPO* transcription [59]. Moreover, genes (lncRNA and mRNA) in the co-expression network significantly overlapped with each other (Fig. 2E) and were also enriched in epidermal development and cell cycle-related biological processes, thus indicating the vital role of lncRNA in the tumorigenesis of both cancer subtypes (Fig. 2F).

Additionally, we identified a number of genes from TCGA that were independently correlated with survival in EAC, and ESCC (Table S1). For example, higher expression of *KRAS* and *CDX2* was associated with poor prognosis in EAC and ESCC (Fig. 3A, B). *CD180* is highly expressed in ESCC and has been reported to be a prognostic marker for radioresistance in ESCC [60]. Additionally, *EFNA1* widely affects tumor growth by enhancing tumor angiogenesis, malignant cell events, and invasiveness in gastrointestinal tumors [61,62], and leads to poor prognosis in ESCC.

3.2. Aberrant promoter activity and alternative promoters in ESCC and EAC

The use of alternative promoters has been associated with patient survival [42]. We identified 1913 upregulated and 1977 downregulated promoter activities in ESCC and 3008 upregulated and 2798 downregulated promoter activities in EAC (Fig. 4A). Considering that alternative promoters impact isoform diversity, differentially expressed isoforms were detected (2432 upregulated and 2386 downregulated in ESCC; 4790 upregulated and 3377 downregulated in EAC) (Fig. 4B). Next, promoters exhibiting differentially expressed isoforms in ESCC and EAC were compared (Fig. 4C). Notably, downregulated promoter activity related genes in ESCC and EAC were also enriched in epidermal development, cornification, and keratinization, and this was consistent with the downregulated gene expression in these two cancer subtypes (Fig. 4D.e). Meanwhile, the promoter activity of more ECM-related genes was upregulated in ESCC (Fig. 4D.a; Figure S2.a). Additionally, alternative promoters of amoebiasis, human papillomavirus infection, measles, influenza A, and pathogenic *Escherichia coli* (*E. coli*) infection-related genes were observed in ESCC and EAC, thus suggesting that the human body respond to foreign parasites or pathogens during the developments of esophageal cancer (Figure S2).

Given the role of promoters in transcriptional regulation, several

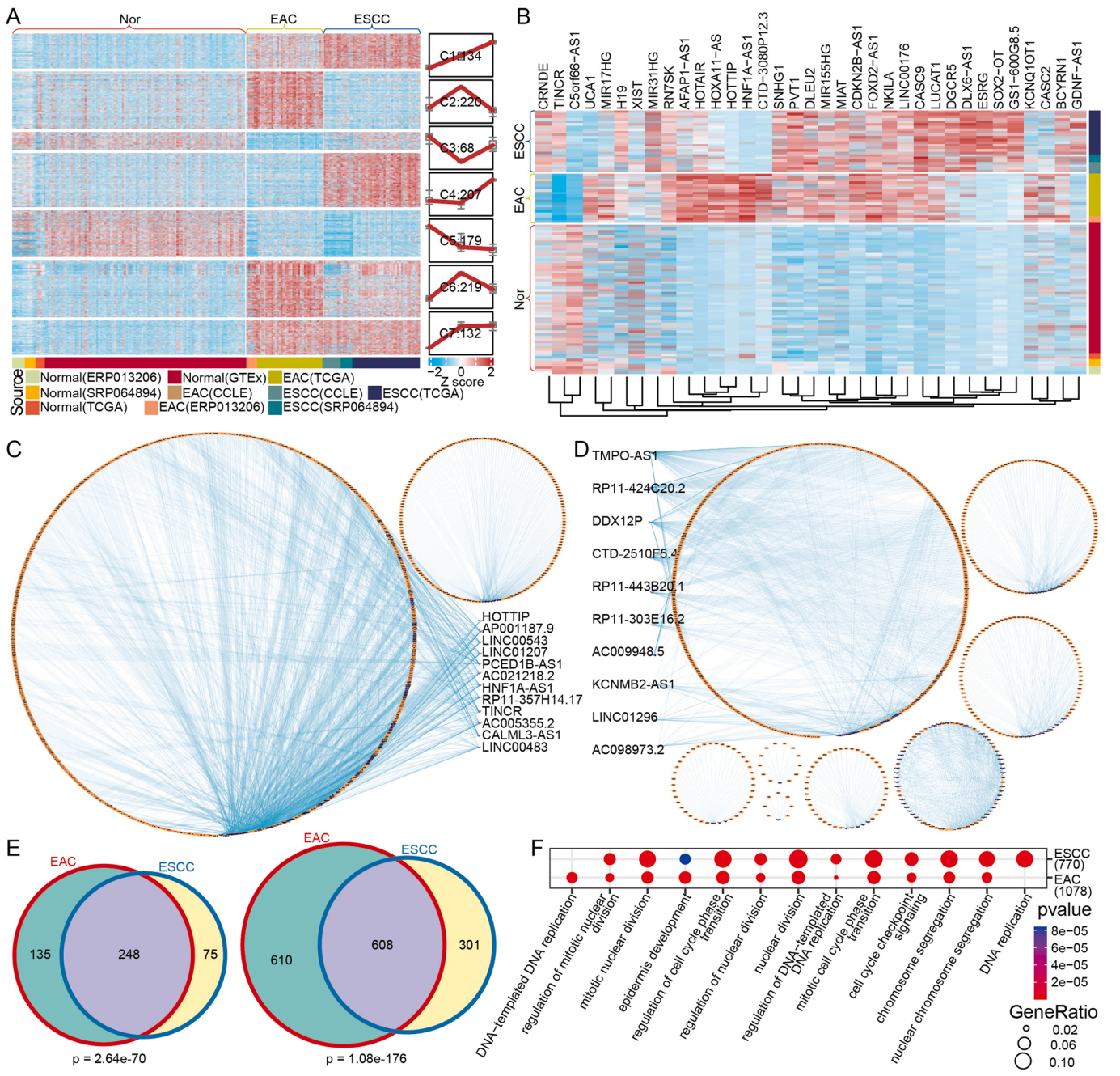

Fig. 2. : Differential lncRNA expression patterns between normal tissue, EAC and ESCC. **A).** Heatmap of differentially expressed lncRNAs in seven clusters. **B).** Gene expression of esophageal cancer-related genes reported in database of lncRNADisease. **C, D).** Co-expression network of differentially expressed lncRNA and mRNA in EAC or ESCC. The "V" shape represents lncRNAs; the Ellipse represents mRNA. The top connected degree lncRNAs or mRNAs were placed in the center. **E).** Overlapped lncRNAs or mRNAs in the co-expression network of EAC and ESCC. **F).** Biological process enrichment of co-expressed mRNAs in the network of EAC and ESCC.

transcription factor motifs (KLF3, KLF5, EHF, SP2, KLF9, EGR2, and KLF4) were observed to be enriched in alternative promoters. (Fig. 4E). Consistent with the decreased promoter activity in ESCC, the mRNA expression of the transcription factors *KLF3*, *KLF5*, and *EHF* were downregulated. In contrast, *SP2* and *KLF9* were upregulated, whereas *KLF4* was downregulated in EAC (Fig. 4E). Abnormal expression of the SP and KLF transcription factors is involved in the tumorigenesis of digestive cancers, including ESCC, EAC, and gastric cancer [63]. The epithelium-specific ETS (ESE) transcription factor EHF has been demonstrated to be a candidate tumor-suppressor in ESCC [64].

In agreement with the transcriptional regulatory role of promoters, most genes with one (86% in ESCC, 90% in EAC) or more (84% in ESCC, 85% in EAC) promoters with differential activity exhibited the same change direction (Fig. 4F). For example, the keratinization related gene, small proline-rich protein 3 (*SPRR3*) down-regulated promoter activity of prmtr.45969 (promoter ID produced by proActiv package), and its gene expression was downregulated in ESCC and EAC (Fig. 4G). E74-like factor 2 (*ELF2*), has been reported to exhibit different functional isoforms during hemopoietic cell development [65]. The aberrant promoter activity of *ELF2* (altered: prmtr.17013 and prmtr.17014; unaltered: prmtr.17010) was selectively present in ESCC with only one differential isoform (ENST00000511006.1) exhibiting this change

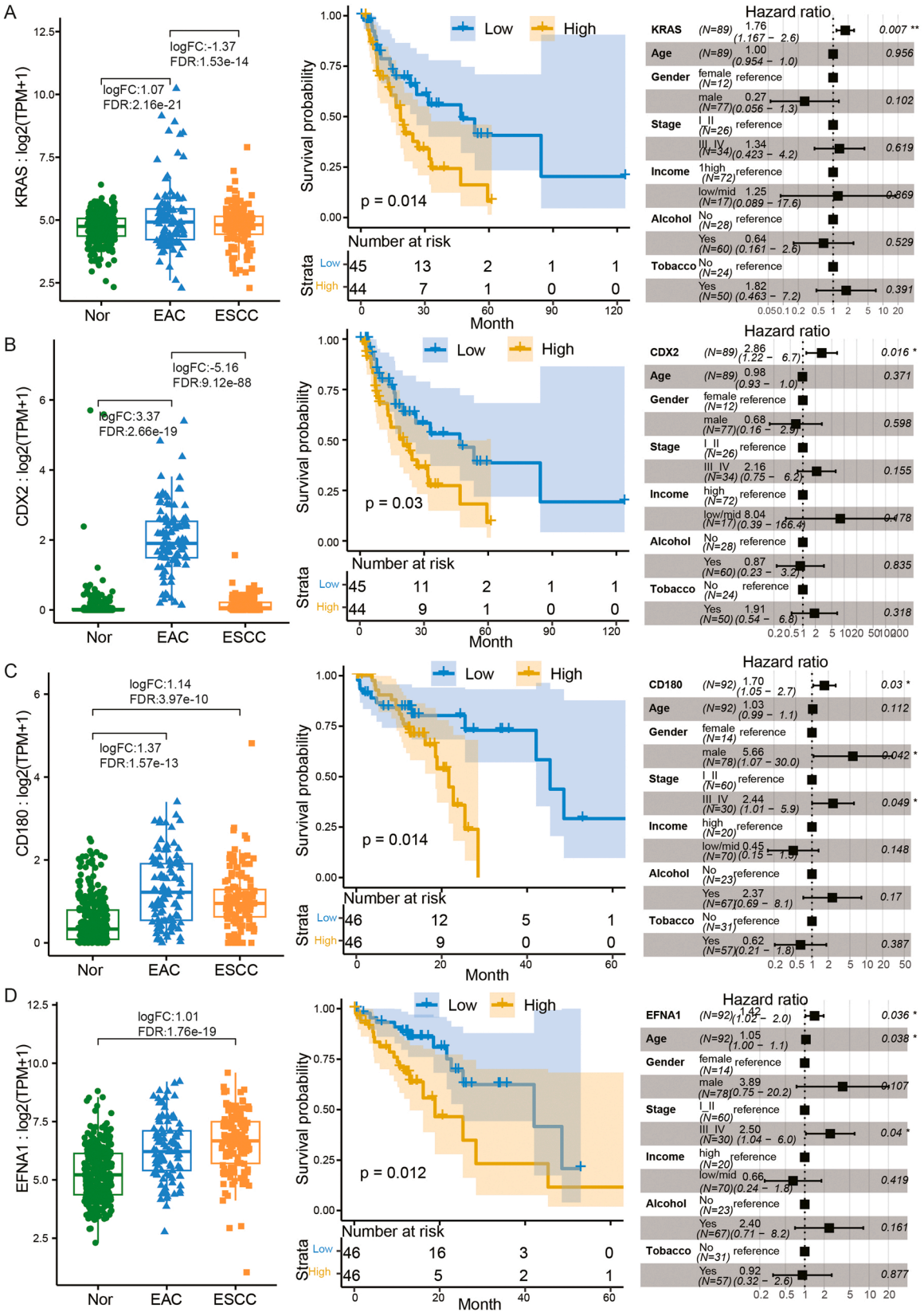

Fig. 3. : Gene expression level, survival curve and Hazard ratio of potential biomarkers for prognosis of EAC (*KRAS* and*CDX2*) and ESCC (*CD180* and *EFNA1*).

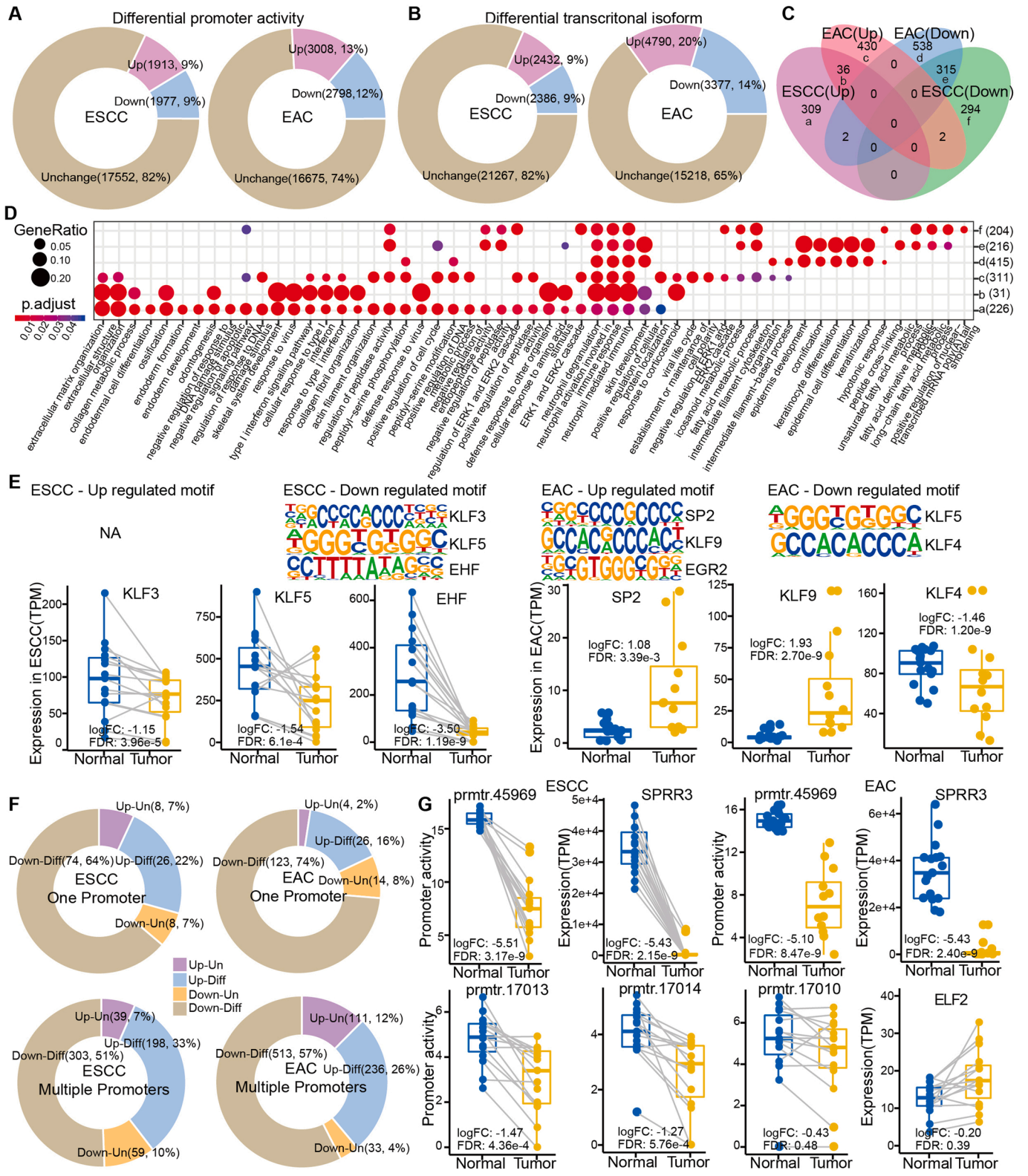

Fig. 4. : Comparison of promoter activity in ESCC and EAC. **A–B).** Distribution of differential promoter activity (**A**) and isoforms (**B**) in ESCC and EAC. **C).** Venn diagram of a promoter with differential activity and yielding differentially expressed transcripts in ESCC and EAC. **D).** Biological process enrichment of genes selected from **C. E).** Motifs of transcriptional factors (TFs) enriched in differential promoter activity regions and the expression level of TFs between ESCC and EAC. **F).** Differential promoter activity (one or more promoters) and expression alterations of their host genes in ESCC and EAC. **G).** Promoter activity and gene expression of *SPRR3* in ESCC and EAC. Alternative promoter activity in *ELF2* with multiple promoters in ESCC.

rather than the total gene expression. This suggests that alternative promoters may play a role in the development of esophageal cancers (Fig. 4G).

3.3. Aberrant alternative splicing in ESCC and EAC

Aberrant RNA alternative splicing (AS) in cancer can act as a driver of tumorigenesis [66,67]. We identified 459 differential AS events between normal and ESCC tissues, including 233 skipped exons (SE), 73 retained intron (RI), 92 mutually exclusive exons (MXE), 27 alternative 5′ splice sites (A5SS), and 34 alternative 3′ splice sites (A3SS) events (Fig. 5A). In contrast, 506 differential AS events were identified between the normal and EAC tissues, including 333 SE, 3 RI, 124 MXE, 22 A5SS, and 24 A3SS events (Fig. 5A). However, only 45 genes with differential AS events were shared between ESCC and EAC samples (Fig. 5B). These genes were enriched in bacterial invasion of epithelial cells and Salmonella infection pathways in ESCC and EAC (Fig. 5C; Figure S3). Additionally, genes related to base excision repair or the Hippo signaling pathway may be affected by AS, particularly in ESCC and EAC. For example, an exon skipping event (chr22:29099493–29099554 located in the S_TK_c and PKc_like domains involved in protein phosphorylation) occurred in the DNA repair gene *CHEK2* in ESCC, and exon retention (chr11:12864249–12865000 located in the TEA domain involved in DNA binding) occurred in the transcription factor *TEAD1* in EAC (Fig. 5D). Furthermore, the majority of the top 10 rMAPS-predicted RBPs involved in AS event formation were present in ESCC and EAC with upregulated and downregulated RBP motif scores, respectively (Fig. 5E). For example, the PCBP2 motif score was enriched upstream of the exon with upregulated AS in ESCC (Fig. 5F). The SFPQ motif score was enriched upstream of the exon with downregulated AS in EAC (Fig. 5F). PCBP2 and SFPQ are reported involved in virus-mediated innate immune responses [68,69].

3.4. Abnormal alternative polyadenylation in ESCC and EAC

APA regulates gene expression by either lengthening or shortening the 3′UTR region. Dysregulation of APA has been linked to various human diseases, such as cancer [70–72]. We identified 608 differential

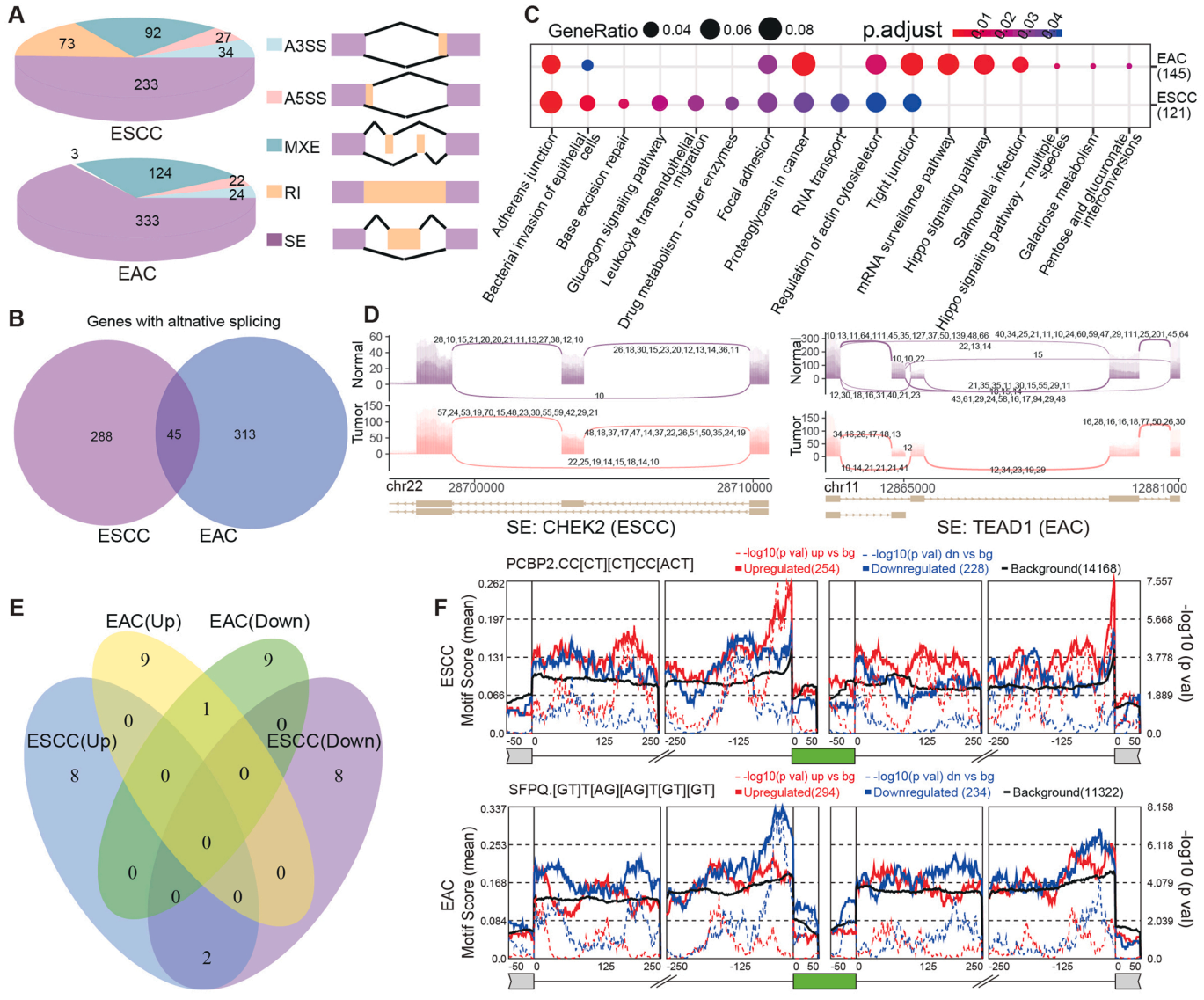

Fig. 5. : Alternative splicing of ESCC and EAC. **A).** Distribution of alternative splicing types in ESCC and EAC. **B).** Overlap of AS events in ESCC and EAC. **C).** Biological process enrichment of AS-related genes in ESCC and EAC. **D).** Visualization of AS in the genes *VCL* and *AFDN*. **E).** Overlap of AS events and differentially expressed genes. **F).** Biological process enrichment of differentially expressed genes with AS events. **G).** Venn diagram of the top 10 enriched RBPs around AS sites. **H).** Distribution of two RBP motif scores in ESCC and EAC.

APAs between normal tissue and ESCC, including 359 lengthened and 249 shortened 3′UTRs (Fig. 6A left). The typical PAS motifs AAUAA and CGKCM were significantly enriched in the dPAS and pPAS regions, respectively (Fig. 6A; Figure S3B). Meanwhile, 1160 differential APAs were identified between normal tissue and EAC, including 826 lengthened and 344 shortened 3′UTRs (Fig. 6A right). Similarly, another PAS site motif, AWUAAA, was enriched around lengthened dPAS regions. Further, a comparison of the APA profiles revealed that most genes related to differential APA were specifically occupied in ESCC and EAC with a shortened or lengthened 3′UTR (Fig. 6B). These genes were enriched in distinct pathways, including the WNT signaling pathway, pathogenic *E. coli* infection, pancreatic cancer, viral carcinogenesis, autophagy, the Hippo signaling pathway, and the cell cycle (Fig. 6C; Figure S4 A). Around the flanking 200 bp of PAS sites (Figure S4C), the top 2 RNA motifs were identified, including AAUAA (−50−0) and UAMA (–75 to –25) in ESCC lengthened 3′UTR, CSC (–100 to –50) and GAGR (–125 to –75) in ESCC shortened 3′UTR, AWUAAA (−50−0) and RUAUW (–175 to –125) in EAC lengthened 3′UTR, and GGAS (–200 to –150), and SRCC (–100 to –50) in EAC shortened 3′UTR (Fig. 6D).

Moreover, genes with shortened or lengthened APA may affect gene expression via RBPs or miRNA binding. We identified 62, 72, 24, and 134 genes with lengthened or shortened APA with upregulated or downregulated gene expression in ESCC (Fig. 6E, left). Additionally, 189, 153, 62, and 209 events were identified in the EAC (Fig. 6E, right). UpSet plot [73] analysis indicated that only a small number of these genes were shared between ESCC and EAC (Fig. 6F). The genes from the eight groups were observed to be enriched in several group-specific KEGG pathways, including human papillomavirus infection, p53 signaling pathway, bacterial invasion of epithelial cells, hepatitis C, pathogenic *E. coli* infection, ECM-receptor interaction, viral myocarditis, Kaposi sarcoma-associated herpes virus infection, the mTOR signaling pathway, and the Hippo signaling pathway (Fig. 6G). Several group-specific biological processes were identified (Figure S4D). Genes that were upregulated with shortened or lengthened 3′UTR were enriched in ECM-related biological processes in ESCC and EAC (Fig. 6E; Figure S4D). Furthermore, potential RBPs or miRNAs may be involved in group-specific lengthened or shortened APA function, including AAUAAH, AAUAAA, AAKAAA matched KHDRBS3 or has-miR-126–5p, CCAGSCUGG matched has-miR-4673, and AGGCWGGAG matched SMAD4A in ESCC or EAC (Figure S4E). Among the 22 core APA RBPs [74,75], polyadenylate-binding protein 1 (PABPC1) was downregulated in ESCC and EAC. Upregulated cleavage and polyadenylation-specific factor 6 (CPSF6) and downregulated protein phosphatase 1 catalytic subunit beta (PPP1CB) were occurred in the EAC (Figure S4F). This suggests that ESCC and EAC possess different APA profiles.

3.5. Gene fusions in ESCC and EAC

Numerous gene fusions have been confirmed to be involved in tumorigenesis and represent potential treatment targets [76,77]. Here, we identified 23 and 25 fusion gene pairs in ESCC and EAC, respectively (Fig. 7 A, B; Table S2). Eight gene fusion pairs from ESCC and six from EAC have been reported in public databases. For example, a fusion pair (*CASC8 -CASC19*), ranked first with 49 junction reads and 36 spanning fragments, was identified in ESCC(Fig. 7 C; Table S2). Both *CASC8* and *CASC19* are susceptible lncRNAs involved in breast, bladder, and colorectal cancers. The fusion pair (*CASC8-CASC19*) can be annotated in both "TCGA_StarF2019" and "CCLE_StarF2019". *CCT2-DYRK2* and *ERBB3–EGFR* were identified in EAC (Fig. 7D). The *CCT2-DYRK2* fusion ranked first with 351 junction reads and 38 spanning fragments and may form a new protein with the Cpn60_TCP1 domain of CCT2 and the kinase domain of DYRK2 (Fig. 7D; Table S2). Dual-specificity tyrosine phosphorylation regulated-kinase 2 (DYRK2) belongs to a family of protein kinases involved in the regulation of the cell cycle, cell proliferation, apoptosis, cellular growth, or development through the phosphorylating of its target genes, such as *TP53*, *TERT*, *MYC*, and *JUN*. Amplification and overexpression of DYRK2 has been reported in the lung and esophagus adenocarcinomas [78]. ERBB3 (HER3) and EGFR (HER1) can form a heterodimeric complex [79]. Both ERBB3 and EGFR, which belong to the epidermal growth factor receptor family, are involved in the tumorigenesis of EAC [79,80]. Furthermore, no fusion genes were observed to be common between ESCC and EAC, thus indicating a specific role for gene fusion formation and their potential function in the development of these two cancer subtypes.

4. Discussion

We comprehensively compared the transcriptome characterization of ESCC and EAC at multiple levels, including aberrant mRNA and lncRNA expression, promoter activity, AS, APA, and gene fusion. First, ESCC and EAC share numerous differentially expressed genes that are enriched in the PI3K-Akt signaling pathway, ECM-related processes activation, and epidermal development repression. Additional ESCC-specific genes were enriched in the activation of the PI3K-Akt pathway and ECM-related processes. Additionally, the promoter activities of a greater number of ECM-related genes were upregulated in ESCC. In the activated ECM-related pathways and biological processes, the immune environment may be altered in ESCC and EAC. Genes belonging to the C6 cluster with higher expression levels in the EAC were enriched in the chemokine signaling pathway and biological processes related to T-cell activation. Furthermore, the prediction of immune cell abundance revealed that several immune cell types were significantly higher or lower in cancer tissues than in normal esophageal tissues. Significant differences in the abundance of a number of immune cell types were also observed between ESCC and EAC. This indicated that both the two cancer subtypes possessed their distinct TMEs. In recent years, there has been a surge in the development of neoadjuvant immunotherapy and chemoradiotherapy along with numerous clinical trials. To improve clinical outcomes with appropriate treatment of the two esophageal cancer subtypes, further investigation is necessary to elucidate the TME and the underlying response or non-response molecular mechanisms [13–15]. Single-cell RNA-Seq (scRNA-Seq) has successfully revealed the components of TME in breast cancer [81]. To decode cell types in ESCC or EAC, scRNA-Seq may provide detailed cell populations of both two subtype cancers in the future. We also observed that the esophageal cancer cell lines from CCLE exhibited lower abundance of immune cells compared to cancer tissues. This suggests that the cultured ESCC and EAC cancer cell lines lose the real immune environment acquired from esophageal cancer tissues of the patient.

Meanwhile, the repression of epidermal development and its related biological processes occurred under several conditions, including gene expression, lncRNA co-expression network, and downregulated promoter activity, thus suggesting that a common feature of epidermal development repression occurs in esophageal cancer development. Although most gene expression profiles were generated by poly A tail-enriched RNA-Seq with detection limitations of lncRNAs, several lncRNAs possessing poly A tails have also been identified in ESCC or EAC. These co-expressed lncRNAs may participate in the cell cycle and epidermal development in EAC and ESCC. To comprehensively reveal the role of lncRNAs in ESCC and EAC, additional total RNA-Seq should be performed, and a higher abundance of lncRNAs should be acquired in the future.

Moreover, the identification of differentially enriched transcription factors from regions exhibiting aberrant promoter activity indicated similar and distinct potential regulation of gene expression at the transcriptional level in ESCC and EAC. Additionally, alternative splicing and APA events in ESCC and EAC exhibited more distinct patterns than those of their shared counterparts, including functional enrichment. These events also involve RBPs. Furthermore, no gene fusion was found to be observed between the two cancer subtypes during gene fusion analysis. Additionally, we observed that aberrant differentially expressed genes, promoter activity, alternative splicing, and APA event-related genes

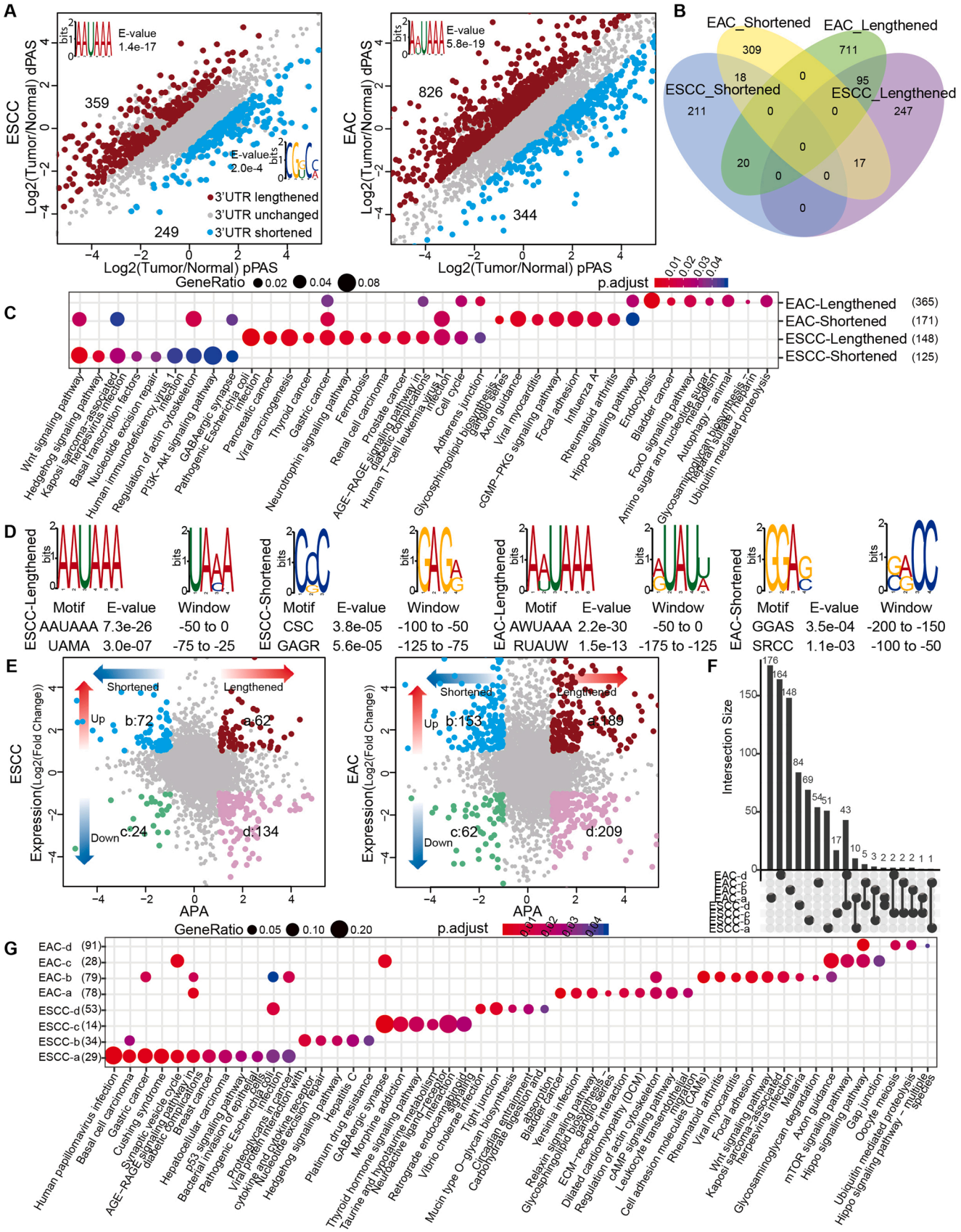

Fig. 6. : Comparison of the alternative polyadenylation (APA) profiles of ESCC and EAC. **A).** The APA profiles of ESCC and EAC. Top enriched motifs of distal and proximal PAS regions. **B).** Venn diagram of genes with shortened and lengthened 3′UTRs in ESCC and EAC. **C).** KEGG pathway enrichment of genes with shortened or lengthened 3′UTRs in ESCC and EAC. **D).** Top two enriched RNA motifs around the shortened or lengthened 3′UTRs in ESCC and EAC. **E).** APA and expression alterations of their host gene in ESCC and EAC. **F).** Upset plot of genes related to **E**. **G).** KEGG pathway enrichment of genes from **E**.

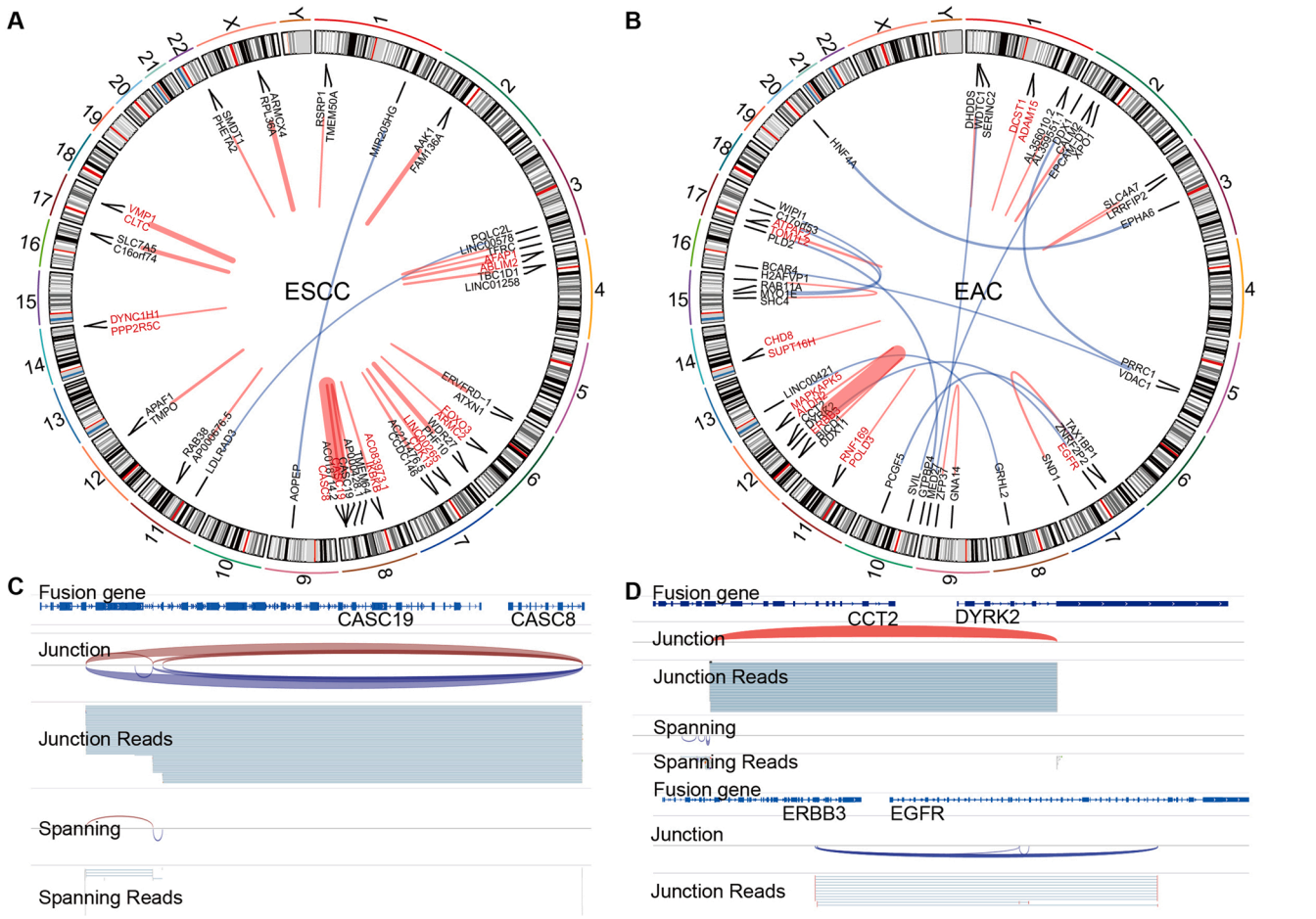

Fig. 7. Gene fusion of ESCC and EAC. **A–B).** Circos plot of gene fusions in ESCC (left) and EAC (right). Gene fusions within the same or different chromosomes were highlighted in red or blue, respectively. A thicker line indicates more reads supporting the connection. **C–D).** Visualization of the gene fusions *CASC19-CASC8* in ESCC and *CCT2-DYRK2* and *ERBB3-EGFR* in EAC. Junction reads and spanning reads from + strand and – strand were visualized by IGV with red and blue, respectively.

were enriched in several viral or bacterial infections in ESCC and EAC. A previous study reported that oral microbiome composition could reflect the prospective risk of esophageal cancers [11,82]. Microbial metabolites inhibit splicing [67]. For example, the Herpes simplex virus (HSV) ICP27 can regulate both the alternative splicing and polyadenylation of pre-mRNA in a sequence-dependent manner [83]. This suggests that certain foreign parasites or pathogens may cause aberrant gene expression and AS and APA events in the context of ESCC and EAC.

5. Conclusions

Therefore, our analysis results offer a comprehensive characterization of multiple shared and distinct signatures in the transcriptomes of ESCC and EAC, including gene expression, promoter activity, alternative splicing, polyadenylation, and gene fusion. This study revealed the potential mechanisms underlying aberrant events and their effects on tumorigenesis in ESCC and EAC. These findings pave the way for the development of effective diagnostic and therapeutic approaches for ESCC and EAC.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

Beijing Municipal Commission of Health and Family Planning Project [PXM2018 026279 000005 to Q.M.Z.]; National Natural Science Foundation of China [81490753, 81502150 to Q.M.Z.]; China Postdoctoral Science Foundation (2018M641117 to X.F.L.); National Natural Science Foundation of China [81902879, 82273431 to X.F.L.].

CRediT authorship contribution statement

Q.M.Z and X.F.L designed the analysis; X.F.L and Q.J.M performed the analysis of transcriptome. Y.W, X.F.L and C.L collected the datasets; K.Z and C.L performed co-expressed analysis and visualization; Q.M.Z, X.F.L and Y.W wrote the manuscript. Q.J.M, W.M.Z, L.S.S and H.J.T, B. W participated to revise the manuscript.

Declaration of Competing Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Luxia Zhang and Yang Xing from the National instate of health data science of Peking University for their helps in maintaining the high performance computing systems. We would like to also thank Editage (www.editage.cn) for English language editing.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.07.030.

References

[1] Smyth EC, et al. Oesophageal cancer. Nat Rev Dis Prim 2017;3:1–21.
[2] Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca Cancer J Clin 2021;71:209–49.
[3] Bray F, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca Cancer J Clin 2018;68:394–424.
[4] Song Y, et al. Identification of genomic alterations in oesophageal squamous cell cancer. Nature 2014;508:91–5.
[5] Torre LA, et al. Global cancer statistics, 2012. CA a Cancer J Clin 2015;65:87–108.
[6] Cui R, et al. Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. Gastroenterology 2009. https://doi.org/10.1053/j.gastro.2009.07.070.
[7] Li XC, et al. A mutational signature associated with alcohol consumption and prognostically significantly mutated driver genes in esophageal squamous cell carcinoma. Ann Oncol 2018;29:938–44.
[8] Chang J, et al. Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. Nat Commun 2017;8:1–11.
[9] Wu C, et al. Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nat Genet 2011;43:679–84.
[10] Kim J, et al. Integrated genomic characterization of oesophageal carcinoma. Nature 2017;541:169–74.
[11] Peters BA, et al. Oral microbiome composition reflects prospective risk for esophageal cancers. Cancer Res 2017;77:6777–87.
[12] Zhu F, et al. Autoreactive T cells and chronic fungal infection drive esophageal carcinogenesis. Cell Host Microbe 2017;21(478–493):e7.
[13] Shah MA, et al. Treatment of locally advanced esophageal carcinoma: ASCO guideline. J Clin Oncol 2020;38:2677–94.
[14] Shah MA, et al. Improving outcomes in patients with oesophageal cancer. Nat Rev Clin Oncol 2023;20.
[15] Kelly RJ, et al. Adjuvant nivolumab in resected esophageal or gastroesophageal junction cancer. N Engl J Med 2021;384:1191–203.
[16] Greally M, Ilson DH. Neoadjuvant therapy for esophageal cancer: who, when, and what? Cancer 2018;124:4276–8.
[17] Teng H, et al. Inter- And intratumor DNA methylation heterogeneity associated with lymph node metastasis and prognosis of esophageal squamous cell carcinoma. Theranostics 2020. https://doi.org/10.7150/thno.42559.
[18] Wang L-D, et al. Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. Nat Genet 2010;42:759–63.
[19] Abnet CC, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nat Genet 2010;42:764–8.
[20] Wu C, et al. Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. Nat Genet 2014;46:1001–6.
[21] Su Z, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. Nat Genet 2012;44:1131–6.
[22] Gharahkhani P, et al. Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. Lancet Oncol 2016;17:1363–73.
[23] Levine DM, et al. A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. Nat Genet 2013;45:1487–93.
[24] Lin D-C, et al. Identification of distinct mutational patterns and new driver genes in oesophageal squamous cell carcinomas and adenocarcinomas. *Gut* gutjnl- 2017:2017–314607. https://doi.org/10.1136/gutjnl-2017-314607.
[25] Li CQ, et al. Integrative analyses of transcriptome sequencing identify novel functional lncRNAs in esophageal squamous cell carcinoma. Oncogenesis 2017;6:1–14.
[26] Maag JLV, et al. Novel aberrations uncovered in Barrett's esophagus and esophageal adenocarcinoma using whole transcriptome sequencing. Mol Cancer Res 2017;15:1558–69.
[27] Lonsdale J, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013. https://doi.org/10.1038/ng.2653.
[28] Leinonen R, Sugawara H, Shumway M. The sequence read archive. Nucleic Acids Res 2011. https://doi.org/10.1093/nar/gkq1019.
[29] Ghandi M, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature 2019. https://doi.org/10.1038/s41586-019-1186-3.
[30] Chang K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013;45:1113–20.
[31] Goldman M, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. bioRxiv 2019. https://doi.org/10.1101/326470.
[32] Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018. https://doi.org/10.1093/bioinformatics/bty560.
[33] Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019. https://doi.org/10.1038/s41587-019-0201-4.
[34] Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30.
[35] Harrow, J., Frankish, A., Gonzalez, J.M. & Frazer, K.A. GENCODE: The reference human genome annotation for The ENCODE Project. 1760–1774 (2012) doi:10.1101/gr.135350.111.
[36] Zhang, Y., Parmigiani, G. & Johnson, W.E. ComBat-Seq: batch effect adjustment for RNA-Seq count data. 1–21 (2020).
[37] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2009. https://doi.org/10.1093/bioinformatics/btp616.
[38] Miao YR, et al. ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. Adv Sci 2020;7.
[39] Shannon P, et al. Cytoscape: a software Environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[40] Kassambara, A. Package ' survminer '. *R* (2018).
[41] Lumley, T. The survival Package. *R News* (2004).
[42] Demircioğlu D, et al. A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. Cell 2019;178(1465–1477):e17.
[43] Shen S, et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci USA 2014. https://doi.org/10.1073/pnas.1419161111.
[44] Garrido-Martín D, Palumbo E, Guigó R, Breschi A. ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. PLoS Comput Biol 2018. https://doi.org/10.1371/journal.pcbi.1006360.
[45] Ha KCH, Blencowe BJ, Morris Q. QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-Seq data. Genome Biol 2018;19:45.
[46] Haas BJ, et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol 2019. https://doi.org/10.1186/s13059-019-1842-9.
[47] Lågstad S, et al. Chimeraviz: A tool for visualizing chimeric RNA. Bioinformatics 2017. https://doi.org/10.1093/bioinformatics/btx329.
[48] Robinson JT, et al. Integrative genomics viewer. Nat Biotechnol 2011. https://doi.org/10.1038/nbt.1754.
[49] Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38:576–89.
[50] Park JW, Jung S, Rouchka EC, Tseng YT, Xing Y. rMAPS: RNA map analysis and plotting server for alternative exon regulation. Nucleic Acids Res 2016;44:W333–8.
[51] Bailey TL. DREME: Motif discovery in transcription factor ChIP-seq data. Bioinformatics 2011. https://doi.org/10.1093/bioinformatics/btr261.
[52] Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol 2007. https://doi.org/10.1186/gb-2007-8-2-r24.
[53] Bailey TL, MacHanick P. Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res 2012. https://doi.org/10.1093/nar/gks433.
[54] Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omi A J Integr Biol 2012. https://doi.org/10.1089/omi.2011.0118.
[55] Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol 2012;196:395–406.
[56] Bao Z, et al. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. Nucleic Acids Res 2019. https://doi.org/10.1093/nar/gky905.
[57] Kretz M, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 2013. https://doi.org/10.1038/nature11661.
[58] Lian Y, et al. HOTTIP: a critical oncogenic long non-coding RNA in human cancers. Mol Biosyst 2016. https://doi.org/10.1039/c6mb00475j.
[59] Luo XJ, et al. LncRNA TMPO-AS1 promotes esophageal squamous cell carcinoma progression by forming biomolecular condensates with FUS and p300 to regulate TMPO transcription. Exp Mol Med 2022;54:834–47.
[60] Wang X-Y, et al. Identification of a three-gene prognostic signature for radioresistant esophageal squamous cell carcinoma. World J Clin Oncol 2023;14:13–26.
[61] Chu L-Y, et al. EFNA1 in gastrointestinal cancer: expression, regulation and clinical significance. World J Gastrointest Oncol 2022;14:973–88.
[62] Jiang H, et al. Targeting EFNA1 suppresses tumor progression via the cMYC-modulated cell cycle and autophagy in esophageal squamous cell carcinoma. Discov Oncol 2023;14.
[63] Kim CK, He P, Bialkowska AB, Yang VW. SP and KLF transcription factors in digestive physiology and diseases. Gastroenterology 2017;152:1845–75.
[64] Wang L, et al. Abnormal localization and tumor suppressor function of epithelial tissue-specific transcription factor ESE3 in esophageal squamous cell carcinoma. PLoS One 2015. https://doi.org/10.1371/journal.pone.0126319.

[65] Guan FHX, et al. The antiproliferative ELF2 isoform, ELF2B, induces apoptosis in vitro and perturbs early lymphocytic development in vivo. J Hematol Oncol 2017; 10:1–17.
[66] Teng H, et al. Transcriptomic signature associated with carcinogenesis and aggressiveness of papillary thyroid carcinoma. Theranostics 2018;8:4345–58.
[67] Escobar-Hoyos L, Knorr K, Abdel-Wahab O. Aberrant RNA splicing in cancer. Annu Rev Cancer Biol 2019;3:167–85.
[68] Morchikh M, et al. HEXIM1 and NEAT1 long non-coding RNA form a multi-subunit complex that regulates DNA-mediated innate immune response. Mol Cell 2017. https://doi.org/10.1016/j.molcel.2017.06.020.
[69] You F, et al. PCBP2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase AIP4. Nat Immunol 2009. https://doi.org/10.1038/ni.1815.
[70] Xia Z, et al. Dynamic analyses of alternative polyadenylation from RNA-Seq reveal a 3′-UTR landscape across seven tumour types. Nat Commun 2014;5:1–13.
[71] Chen M, et al. 3′ UTR lengthening as a novel mechanism in regulating cellular senescence. Genome Res 2018:1–10. https://doi.org/10.1101/gr.224451.117.
[72] Xiang Y, et al. Comprehensive characterization of alternative polyadenylation in human cancer. JNCI J Natl Cancer Inst 2017;110:1–11.
[73] Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph 2014. https://doi.org/10.1109/TVCG.2014.2346248.
[74] Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet 2013;14:496–506.
[75] Neve J, Patel R, Wang Z, Louey A, Furger AM. Cleavage and polyadenylation: Ending the message expands gene regulation. RNA Biol 2017;14:865–90.
[76] Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer 2015;15:371–81.
[77] Gao Q, et al. Driver fusions and their implications in the development and treatment of human cancers. Cell Rep 2018;23(227–238):e3.
[78] Miller CT, et al. Amplification and overexpression of the dual-specificity tyrosine-(Y)phosphorylation regulated kinase 2 (DYRK2) gene in esophageal and lung adenocarcinomas. Cancer Res 2003;63:4136–43.
[79] Fichter CD, et al. EGFR, HER2 and HER3 dimerization patterns guide targeted inhibition in two histotypes of esophageal cancer. Int J Cancer 2014;135:1517–30.
[80] Ebbing EA, et al. Esophageal Adenocarcinoma Cells and Xenograft Tumors Exposed to Erb-b2 Receptor Tyrosine Kinase 2 and 3 Inhibitors Activate Transforming Growth Factor Beta Signaling, Which Induces Epithelial to Mesenchymal Transition. Gastroenterology 2017;153(63–76):e14.
[81] Azizi E, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. Cell 2018;174(1293–1308):e36.
[82] Zhu F, et al. Autoreactive T cells and chronic fungal infection drive esophageal carcinogenesis. Cell Host Microbe 2017;21(478–493):e7.
[83] Tang S, Patel A, Krause PR. Herpes simplex virus ICP27 regulates alternative pre-mRNA polyadenylation and splicing in a sequence-dependent manner. Proc Natl Acad Sci 2016;113:12256–61.