



Extrachromosomal circular DNAs are common and functional in esophageal squamous cell carcinoma

Zhenguo Sun¹, Na Ji², Renchang Zhao¹, Jinghui Liang¹, Jin Jiang¹, Hui Tian¹

¹Department of Thoracic Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China; ²Medical Division, Shandong Provincial Western Hospital, Shandong Provincial Ear Nose Throat Hospital, Jinan, China

Contributions: (I) Conception and design: Z Sun; (II) Administrative support: Z Sun, H Tian; (III) Provision of study materials or patients: R Zhao, J Liang, J Jiang; (IV) Collection and assembly of data: Z Sun; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Hui Tian; Zhenguo Sun. Department of Thoracic Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan 250012, China. Email: tianhuiql2010@163.com; sunzg@sdu.edu.cn.

Background: Esophageal squamous cell carcinoma (ESCC) is the leading cause of cancer-related mortality. While recent studies have documented the presence of extrachromosomal circular DNAs (eccDNAs) in various tumors, to date, there have been no studies examining the distribution and function of eccDNAs in ESCC.

Methods: The eccDNAs from three surgically matched ESCC tissue samples were extracted and amplified by rolling circle amplification after removal of linear DNA and mitochondrial circular DNA. High-throughput eccDNA sequencing and bioinformatics analysis was performed to study the distribution pattern and the level of eccDNA expression. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed on the genes associated with the differentially expressed eccDNAs. Five up-regulated and five down-regulated candidate eccDNAs were validated by routine polymerase chain reaction (PCR), TOPO-TA cloning and Sanger sequencing. The nucleotides flanking the eccDNA junctions were analyzed to explore the mechanisms of eccDNA formation.

Results: A total of 184,557 eccDNAs was identified. The overall length distribution ranged from 33 to 968,842 base pairs (bp), with the peak at approximately 360 bp. These eccDNAs mainly originated from 5'- and 3'-untranslated regions (UTRs), and rarely from exons, introns, LINE, or Alu repeat regions. The chromosome distribution, length distribution, and genomic annotation of the eccDNAs were comparable between ESCC samples and matched normal epithelium. Nevertheless, 16,031 eccDNAs were found to be differentially expressed between ESCC and matched normal epithelium, including 10,126 up-regulated eccDNAs and 5,905 down-regulated eccDNAs. GO analysis and KEGG pathway analysis showed enriched in cancer pathways, mitogen-activated protein kinase (MAPK) pathway, GTPase-related activity, and cytoskeleton function. PCR, TOPO-TA cloning, and Sanger sequencing validated the junctional sites of five up-regulated candidate eccDNAs and four other unexpected eccDNAs. A repeat nucleotide pattern between the position flanking the start site and that flanking the end site was detected.

Conclusions: This study demonstrated the genome-wide presence of eccDNAs, explored the differential expression of eccDNAs, and revealed the potential mechanisms of eccDNAs in ESCC. This work provides further insights into our understanding of genome plasticity, the role of eccDNAs in ESCC, and may contribute to the development of potential clinical therapies.

Keywords: Extrachromosomal circular DNAs (eccDNAs); microDNA; esophageal squamous cell carcinoma (ESCC); tumor progression; tumor induction; high-throughput sequencing; differential expression

Submitted Jul 30, 2021. Accepted for publication Sep 09, 2021.

doi: 10.21037/atm-21-4372

View this article at: <https://dx.doi.org/10.21037/atm-21-4372>

Introduction

Esophageal cancer ranks seventh in the incidence of cancers and sixth in cancer-related mortalities worldwide (1). Esophageal squamous cell carcinoma (ESCC) accounts for 90% of all esophageal malignancies globally and remains the dominant pathological type of esophageal cancer in high-risk areas such as China (2). Despite rapid progress in the diagnosis and treatment of ESCC, the overall 5-year survival rate remains poor (3). Unfortunately, to date, the molecular mechanisms underlying the high incidence and poor prognosis associated with ESCC are yet to be fully elucidated.

The majority of human cellular DNA is contained in 22 linear autosome pairs and a pair of sex-determining chromosomes. Extrachromosomal DNA are DNA elements that have separated from the chromosomes. The major form of extrachromosomal DNA is extrachromosomal circular DNA (eccDNA) characterized by a closed circular structure (4). EccDNA was discovered half a century ago when double minutes (DMs) were detected in malignant tumor specimens from children (5). The size of eccDNAs varies from hundreds of base pairs (bp) to several megabases (Mb), however, the majority of eccDNAs are short, ranging from 200–400 bp, and are thus called microDNA or small poly-disperse circular DNA (spcDNA) (4,6). In common physiology, the presence of eccDNAs is linked to normal functions and certain phenotypes, and may involve in genome's plasticity and evolution (7). More recently, it has been shown that eccDNAs could participate in tumorigenesis and tumor progression. The eccDNAs with longer sizes, such as DMs, participate in cancer initiation by carrying oncogenes and promoting gene amplification (8,9). However, less is known about the function and mechanism of the more widespread, smaller type of eccDNAs, as they are too small to contain coding genes (10). Recently, high-throughput sequencing demonstrated that the size distribution of eccDNAs varied between maternal and fetal plasma (11). Furthermore, the genomic annotation of the eccDNA sequences weakly distinguished prostate from ovarian cancer cells (12). The eccDNAs identified in lung cancer samples were longer than that in matched normal tissue and the size of circulating eccDNAs in the plasma decreased after surgical resection of lung cancer (10). These studies suggested that the distribution of eccDNAs may be correlated to tumor cell lineage and the formation of certain malignancies. Moreover, tens of thousands of eccDNAs from smaller ones to longer ones carrying known

cancer drive genes were identified by pan-cancer analysis from various tumor types, including ESCC (13), however, it mainly focused on tumor cell lines or tissue samples, lacking of the comparison of the distribution pattern of eccDNAs in ESCC and matched normal esophageal epithelium. To the best of our knowledge, this is the first report to investigate the distribution, function, and mechanism of eccDNAs in matched ESCC samples. We present the following article in accordance with the MDAR reporting checklist (available at <https://dx.doi.org/10.21037/atm-21-4372>).

Methods

Tissue specimens

This study was approved by the Ethics Committee at Qilu Hospital, Cheeloo College of Medicine, Shandong University, China [reference number KYLL-2021(KS)-013], in accordance with the ethical guidelines of the Declaration of Helsinki (as revised in 2013). Written informed consents were obtained from three male patients (aged 56, 62, and 64 years) with ESCC undergoing surgery in the Department of Thoracic Surgery at Qilu Hospital. Three pairs of tumor tissue samples and matched non-tumor esophageal epithelium samples were collected from specimens obtained during surgery.

Tissue DNA preparation and eccDNA sequencing

High-throughput eccDNA sequencing and subsequent bioinformatics analyses were performed by CloudSeq Biotech Inc. (Shanghai, China). Specially, three pairs of tumor and normal epithelium tissues were suspended in L1 solution (Plasmid Mini AX; A&A Biotechnology, Pomorskie Gdynia, Poland) and supplemented with Proteinase K (ThermoFisher, MA, USA) before incubation overnight at 50 °C with agitation. After lysis, samples were treated with alkaline, followed by precipitation of proteins and separation of chromosomal DNA from circular DNA through an ion exchange membrane column (Plasmid Mini AX; A&A Biotechnology). Column-purified DNA was treated with FastDigest MssI (ThermoFisher, MA, USA) to remove mitochondrial circular DNA and incubated at 37 °C for 16 hours. The remaining linear DNA was removed by exonuclease (Plasmid-Safe ATP-dependent DNase, Epicentre, Wisconsin, USA) at 37 °C in a heating block and the enzyme reaction was carried

out continuously for 1 week, with application of additional ATP and DNase every 24 hours (30 units per day) according to the manufacturer's protocol (Plasmid-Safe ATP-dependent DNase, Epicentre). EccDNA-enriched samples were used as templates for phi29 polymerase amplification reactions (REPLI-g Midi Kit). EccDNA was amplified at 30 °C for 2 days (46–48 hours). Phi29-amplified DNA was sheared by sonication (Diagenode Bioruptor, NJ, Belgium), and the fragmented DNA was subjected to library preparation with NEBNext® Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, MA, USA). Sequencing was carried out on Illumina NovaSeq 6000 with 150 bp paired end mode according to the manufacturer's instructions.

Sequencing analysis of eccDNA

Paired-end reads were harvested from the Illumina NovaSeq 6000 sequencer and quality-controlled by Q30. After 3' adaptor-trimming and removal of low-quality reads by cutadapt software (v1.9.1), the high-quality clean reads were aligned to the reference genome (UCSC hg19) with bwa software v (v0.7.12). The circle-map software (v1.1.4) was used to detect eccDNA within all samples, and samtools (v0.2) software was used to obtain raw soft-clipped read counts of the break point. The edgeR (v0.6.9) software was used to perform normalization and differentially expressed eccDNAs were filtered with a P value <0.05 and fold change (>2 or <1/2). Bedtools (v2.27.1) software was used to annotate the eccDNAs. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed based on the differentially expressed eccDNA-associated genes. IGV (v2.4.10) software was used for eccDNA visualization.

Motif analysis of eccDNA junctional sites

The nucleotide pattern flanking the eccDNA junctional sites was analyzed to explore the mechanism of generation of these eccDNAs. Specially, the nucleotide composition from 10 bp upstream to 10 bp downstream of the start and end position inferred from the reference genome (hg19) was analyzed for each eccDNA locus.

Validation of eccDNA by routine polymerase chain reaction (PCR), TOPO-TA cloning, and Sanger sequencing

Experimental validation was performed on five up-

regulated and five down-regulated eccDNAs in different genomic regions and different chromosomes. Briefly, DNA was extracted from three matched ESCC specimens and treated with FastDigest MssI and exonuclease to remove mitochondrial circular DNA and linear DNA as described above. Rolling circle amplification was performed to increase the yield. PCR was performed using Accurate Taq Master Mix (dye plus) (Accurate Biotechnology, China) to assess the expression levels of the candidate eccDNAs. The reaction conditions were 94 °C for 30 seconds, 28 cycles at 98 °C for 10 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute, followed by final elongation at 72 °C for 2 minutes and storage at 4 °C. The primers of the eccDNAs were designed using the “out-facing” strategy and are described in [Table S1](#). PCR products were loaded onto 1.5% agarose gels and visualized under an ultraviolet Luminescent Image Analyzer (LAS-4000 Mini; GE Healthcare Life Sciences, Pittsburgh, USA). The PCR product with specific positive bands were purified and amplified with TOPO-TA cloning (Zero TOPO-TA Cloning Kit, Shanghai Yeasan, China) and sent for Sanger sequencing (Shanghai Sangon Biotech, China). A comparison of the nucleotide composition of each positive PCR product between Sanger sequencing and high-throughput sequencing was performed.

Statistical analysis

The correlation between the ratio of coding genes/Mb and eccDNA/Mb in any of the chromosomes was performed by linear regression analysis. The data from two groups were compared by Wilcoxon rank-sum test. P value <0.05 was considered to be statistically significant. All statistical analyses were performed using GraphPad Prism 8.0 (GraphPad Software, San Diego, CA, USA).

Results

Genome-wide detection and analysis of eccDNAs in the matched ESCC tissue

High throughput sequencing revealed more than 100 million clean reads in each esophageal epithelium sample. By mapping these clean reads to the human genome (UCSC hg19), 184,557 eccDNAs annotated to 23 pairs of chromosomes were identified. Most of these eccDNAs were detected in more than one specimen ([Table S2](#)). These results indicated that the presence of eccDNAs was a common event in ESCC tissues.

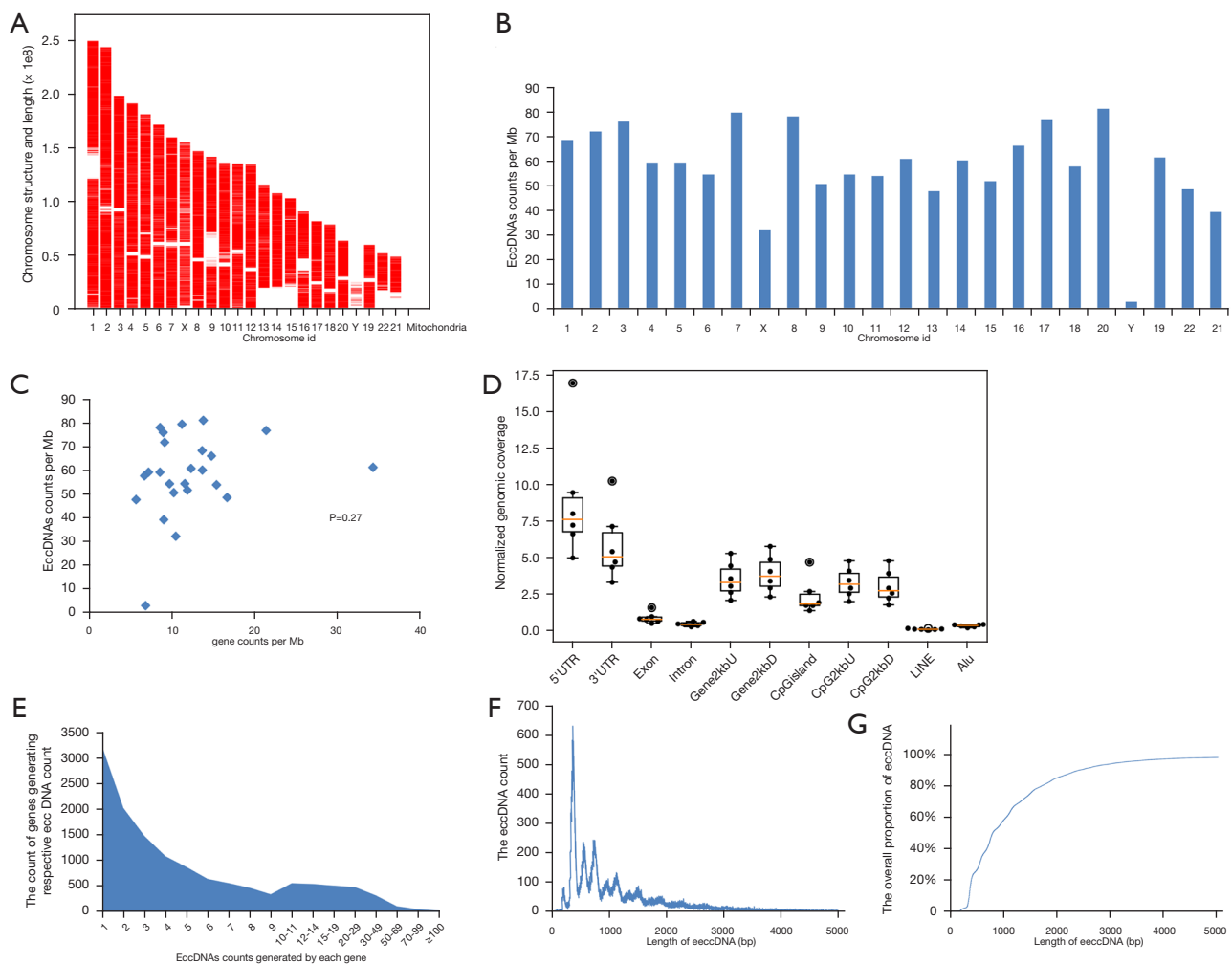


Figure 1 Genome-wide detection and analysis of eccDNA distribution by high-throughput sequencing in matched ESCC samples. (A) The distribution of eccDNAs in the 23 pairs of chromosomes. (B) The eccDNA frequency per Mb in each chromosome. (C) There was no significant relationship between the ratio of coding genes/Mb and eccDNAs/Mb in each chromosome ($P=0.27$). (D) The distribution of eccDNAs in different classes of genomic regions. (E) The distribution of the counts of eccDNAs generated by different genes. (F,G) The length distribution of eccDNA at $\leq 5,000$ bp. EccDNA, extrachromosomal circular DNA; ESCC, esophageal squamous cell carcinoma; Mb, megabase; bp, base pair.

The genomic distribution of the eccDNAs revealed that they were common in each of the 23 pairs of chromosomes. No eccDNAs from the mitochondrion were detected as they were removed before sequencing. The eccDNA frequency per Mb was comparable in each chromosome, except for chromosome Y with a much lower frequency of eccDNAs (Figure 1A,1B). There was no correlation between the ratio of coding genes/Mb and eccDNA/Mb in any of the chromosomes ($P=0.27$; Figure 1C).

All the eccDNAs were mapped to different classes

of genomic regions. Normalized genomic coverage was defined as the percentage of eccDNA mapped to that class of genomic regions divided by the percentage of the genome covered by that class of genomic region (11). The eccDNAs mainly originated from the 5'- and 3'-untranslated regions (UTRs), 2 kb upstream or downstream of genes, and 2 kb upstream to 2 kb downstream of CpG island regions. However, eccDNAs were rarely distributed in exons, introns, LINE, or Alu repeat regions (Figure 1D).

All the eccDNAs identified were derived from 13,177

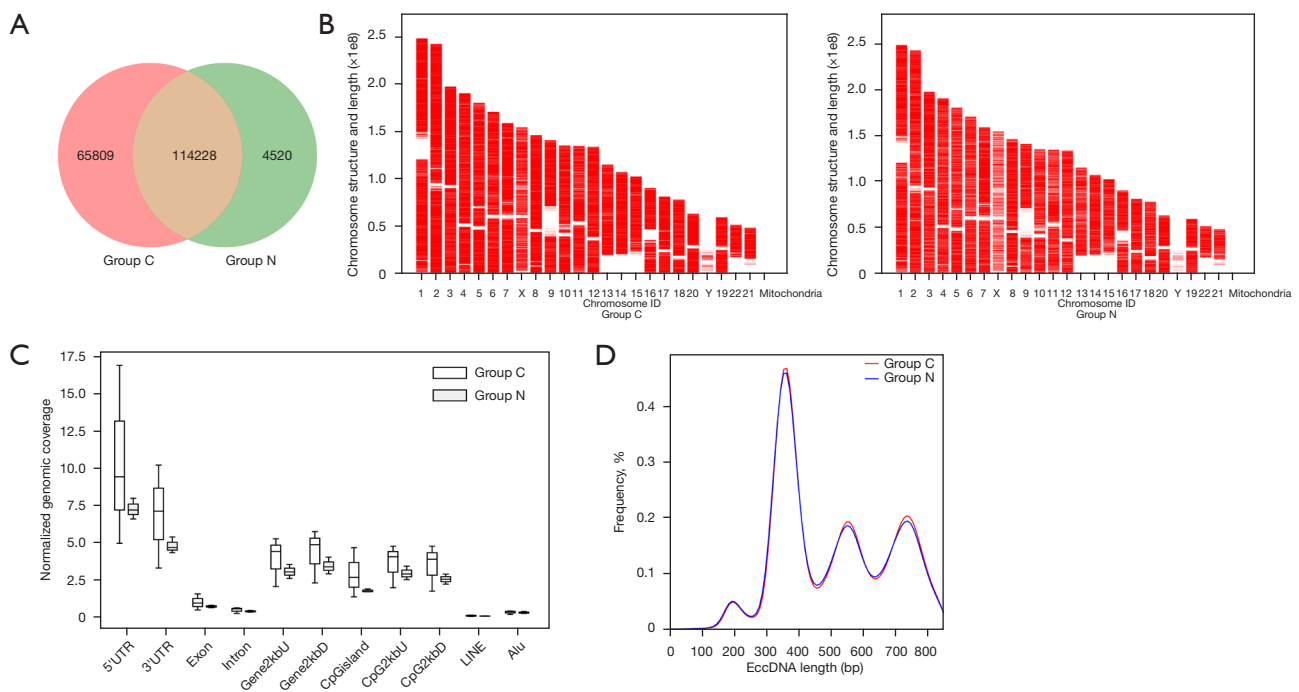


Figure 2 A comparison of the distribution pattern of the eccDNAs between ESCC and matched normal epithelium. (A) A Venn diagram showing the eccDNAs detected in the ESCC and matched normal epithelium samples. (B-D) The chromosome distribution, annotation of genomic elements, and length distribution of eccDNAs were comparable in ESCC and matched normal epithelium. EccDNA, extrachromosomal circular DNA; ESCC, esophageal squamous cell carcinoma.

genes in the genome. The limbic system associated membrane protein (LSAMP) gene was associated with 192 eccDNAs. More than 100 eccDNAs were derived from 16 genes (Table S3) and only 1 eccDNA originated from 3,208 genes (Figure 1E).

The overall length distribution of the eccDNAs ranged from 33 to 968,842 bp, with a peak at approximately 360 bp. Meanwhile, there were two additional peaks at approximately 555 and 736 bp (Figure 1F). In total, 95.0% (175,363/184,557) of eccDNAs were shorter than 3,000 bp and 86.1% (158,850/184,557) were shorter than 2,000 bp (Figure 1F,1G).

Comparing the distribution pattern of eccDNAs between ESCC and matched normal epithelium samples

The distribution features varied between ESCC samples and matched normal esophageal epithelium. The Venn diagram shows that out of 184,557 eccDNAs, 65,809 eccDNAs were only detected in ESCC samples, 4,520 were only detected in normal esophageal epithelium, while 114,228 eccDNAs were detected in both samples

(Figure 2A). However, the chromosome distribution and annotation of genomic elements of the eccDNAs was similar between ESCC and matched normal epithelium (Figure 2B,2C).

Previous studies have demonstrated that the length distribution of eccDNAs is different between maternal and fetal plasma, and between lung cancer tissues and normal lung tissues (10,11). Thus, the length distribution of eccDNAs in the ESCC samples and the matched normal esophageal epithelium was examined. The length distribution in the ESCC and normal esophageal epithelium showed similar features, including the location of the peak and the span of the length of eccDNAs (Figure 2D).

Identification of the differentially expressed eccDNAs between ESCC and matched normal epithelium

Most of the eccDNAs were detected in more than one specimen. Thus, the level of eccDNA expression in ESCC and normal esophageal epithelium was compared. According to the screening criteria (P value <0.05 and |LogFC| >1), a total of 16,031 eccDNAs was defined

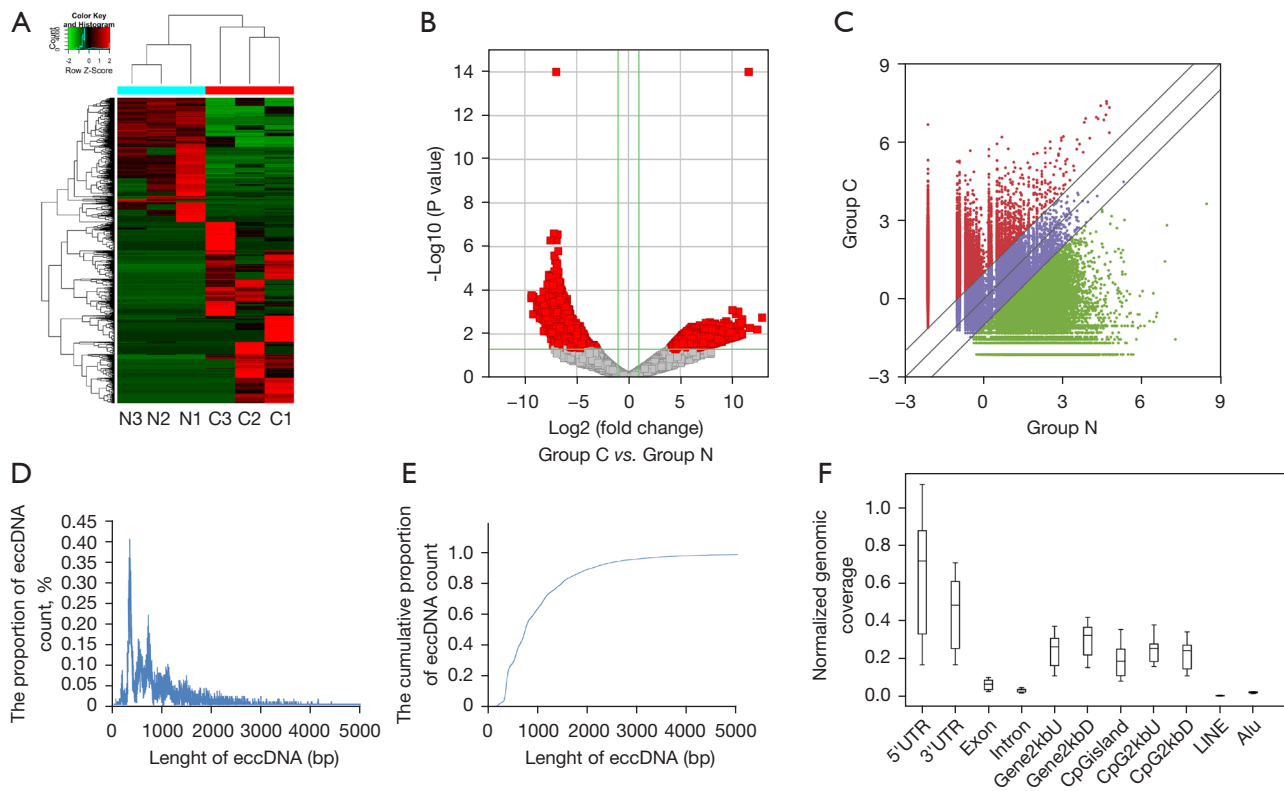


Figure 3 Identification of the differentially expressed eccDNAs between ESCC and matched normal epithelium. (A-C) Cluster, scatter plot, and volcano plot showing the differentially expressed eccDNAs between ESCC and matched normal epithelium. (D,E) The length distribution of the differentially expressed candidate eccDNAs. (F) The distribution of the differentially expressed candidate eccDNA in different classes of genomic regions. EccDNA, extrachromosomal circular DNA; ESCC, esophageal squamous cell carcinoma.

as candidate functional eccDNAs, including 10,126 up-regulated eccDNAs and 5,905 down-regulated eccDNAs. Most of these candidate eccDNAs were detected in either the ESCC samples or the normal esophageal epithelium, while only a small fraction of candidate eccDNAs were detected in both ESCC and normal samples (Figure 3A-3C). These candidate eccDNAs may be potential biomarkers to distinguish ESCC from normal esophageal epithelium and may participate in the initiation and progression of ESCC.

The length distribution of these candidate eccDNAs ranged from 44 to 395,264 bp, with a peak at approximately 357 bp and two additional peaks at 549 and 733 bp (Figure 3D,3E). Mapping of these candidate eccDNAs to genomic elements revealed that they were mainly derived from the 5'- and 3'-UTRs, and rarely from exons, introns, or repeat regions such as LINE and Alu (Figure 3F). Specially, among the 10,126 up-regulated eccDNAs, 49.4% (5,007/10,126) were derived from the region of the genes, while 50.6% (5,119/10,126) were

derived from the intergenic regions. These eccDNAs originated from 3,219 genes. Most of the eccDNAs were derived from the activator of transcription and developmental regulator 2 (AUTS2) gene. The top 10 genes giving rise to the up-regulated eccDNAs are shown in Table S4. Among the 5,905 down-regulated eccDNAs, 48.4% (2,859) and 51.6% (3,046) were derived from the region of genes and intergenic regions, respectively. The down-regulated eccDNAs were derived from 2,235 genes, with the LSAMP, CUB and sushi domain-containing protein 1 (CSMD1) and bicaudal D homolog 1 (BICD1) genes giving rise to the most eccDNAs. The top 10 genes giving rise to the down-regulated eccDNAs are shown in Table S4.

GO and KEGG pathway analysis of genes associated with the differentially expressed eccDNAs

To examine the functions of the genes associated with

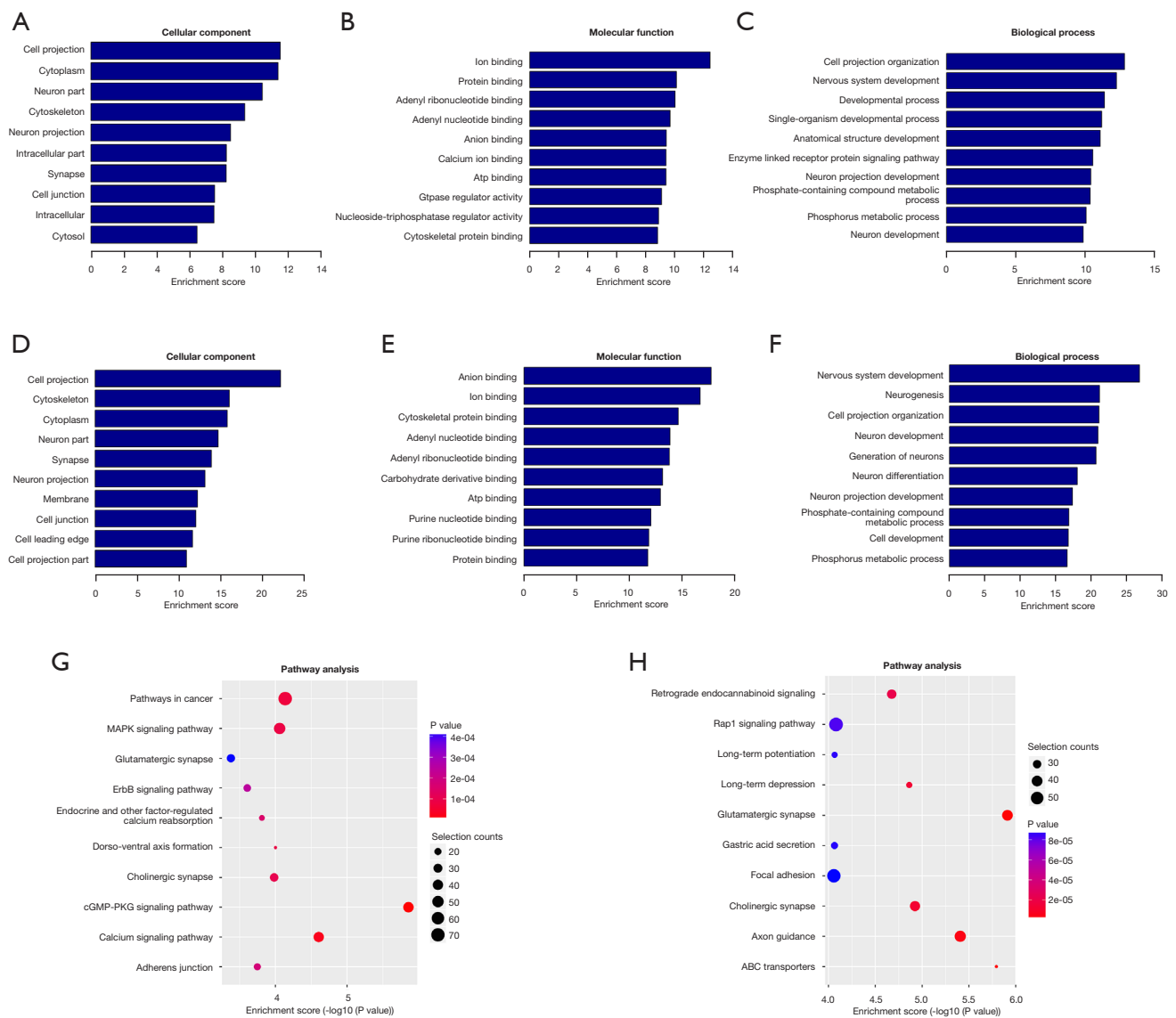
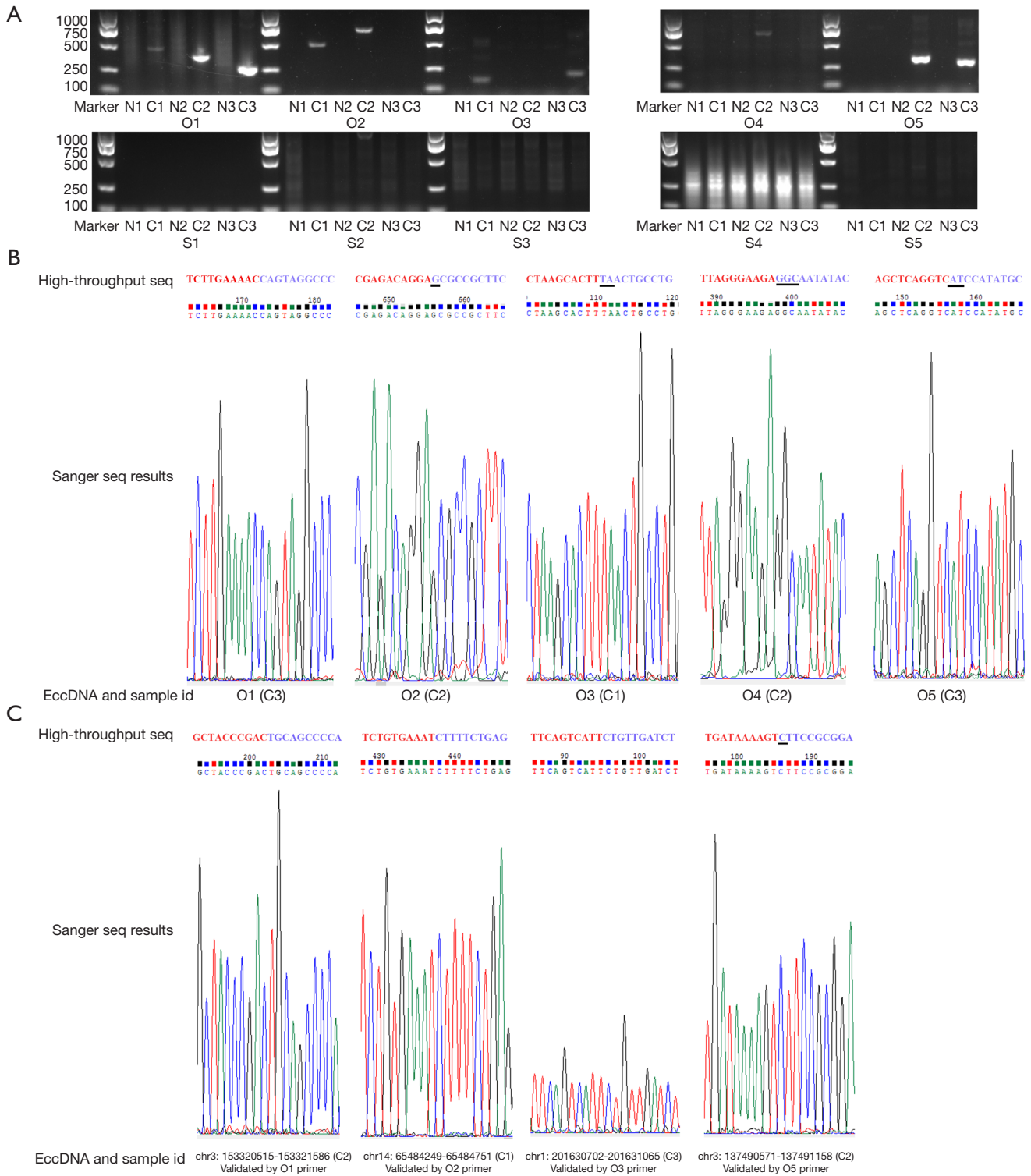


Figure 4 GO and KEGG pathway analysis of the genes associated with the differentially expressed eccDNAs. (A-C) The cellular components, molecular function, and biological processes associated with the down-regulated eccDNAs. (D-F) The cellular components, molecular function, and biological processes associated with the up-regulated eccDNAs. (G) KEGG pathway analysis of the down-regulated eccDNAs in matched ESCC samples. (H) KEGG pathway analysis of the up-regulated eccDNAs in matched ESCC samples. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; eccDNA, extrachromosomal circular DNA; ESCC, esophageal squamous cell carcinoma.

the differentially expressed eccDNAs, GO analysis was performed, including identification of the related cellular components, molecular functions, and biological processes (Figure 4A-4F). The dominant biological processes were related to neurons, and the main molecular function and cellular component was GTPase-related activity and the

cytoskeleton, respectively. In addition, KEGG pathway analysis revealed that the genes associated with the differentially expressed eccDNAs were mainly involved in cancer-related pathways, the mitogen-activated protein kinase (MAPK) pathway, focal adhesion, and the Rap1 pathway (Figure 4G, 4H).



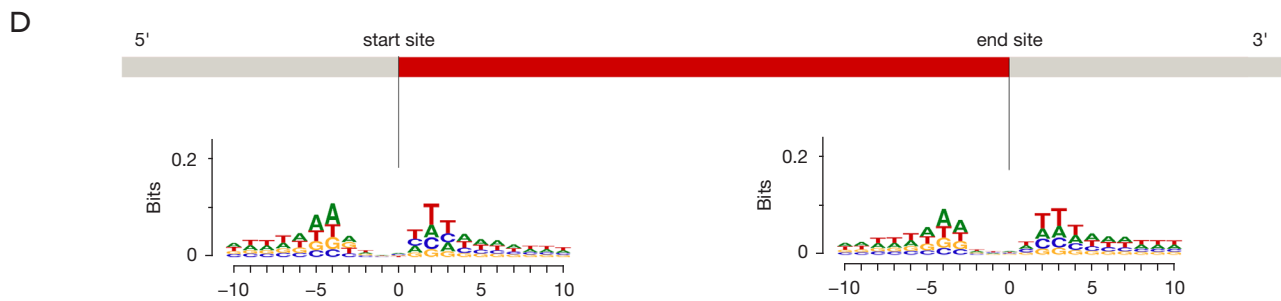


Figure 5 Validation of the junctional sites of the eccDNAs and motif analysis of the flanking junctional sites. (A) The PCR results of the five down-regulated and five up-regulated candidate eccDNAs in three pairs of matched ESCC samples. (B) The results of Sanger sequencing validated the junctional site of the five up-regulated eccDNA in tumor samples. In high-throughput sequencing, red and blue color indicate the two ends of the junctional sites, while the nucleotides with underscore indicate identical nucleotides at both ends. (C) The results of Sanger sequencing unexpectedly validated the junctional sites of the other four true eccDNAs detected by high-throughput sequencing. (D) The motif analysis of the flanking junctional sites by high throughput sequencing. EccDNA, extrachromosomal circular DNA; PCR, polymerase chain reaction; ESCC, esophageal squamous cell carcinoma.

Validation of the eccDNAs in the ESCC samples

PCR was performed to verify five up-regulated candidate eccDNAs (designated as ‘Ox’) and five down-regulated candidate eccDNAs (designated as ‘Sx’). Surprisingly, all five up-regulated candidate eccDNAs were strongly expressed in at least one tumor sample and negative in all the normal epithelium sample. Meanwhile, the five down-regulated candidate eccDNAs were negative in all tumor samples and all normal epithelium samples (*Figure 5A*). To further explore the unexpected results observed with the down-regulated candidate eccDNAs, the data from the high-throughput sequencing was compared with the PCR results. Interestingly, only candidate eccDNAs with more split read counts in the high-throughput sequencing (≥ 20) were positive in the PCR results, while those candidate eccDNAs with fewer split read counts in the high-throughput sequencing (≤ 14) were negative in the PCR results (*Table 1*).

TOPO-TA cloning and Sanger sequencing was performed to validate the junctional sites. Interestingly, the junctional sites connecting the start and end site of each up-regulated candidate eccDNA were completely validated in at least one PCR positive product (*Figure 5B*). When the start site and genomic downstream of the end site of a candidate eccDNA, such as O2, O3, O4 or O5, had one or more identical nucleotides, the end site indicated by high-throughput sequencing would be extended to one or more nucleotides downstream of the true end sites because high-throughput sequencing could not determine whether these

identical nucleotides at the junctional site were from the start site or the end site.

Unexpectedly, the junctional site of candidate eccDNAs could not be validated in the other four PCR positive products by Sanger sequencing. To further explore this, these PCR positive products were re-evaluated and found to be slightly longer or shorter than the expected products. When comparing these PCR products to all the eccDNAs detected by high-throughput sequencing, these four PCR products were determined to be identical to the four eccDNAs partially overlapped with our expected eccDNAs (*Figure 5C*). When carefully comparing the data from the high-throughput sequencing to the PCR results obtained by Sanger sequencing, these four unexpected overlapped eccDNAs and the candidate eccDNAs actually coexisted in abundance in those samples. However, the primer pairs that can potentially amplify both eccDNAs could only amplify the unexpected eccDNAs in the actual PCR reaction due to unknown factors, such as competitive binding to primer pairs or amplifying inefficiency. Thus, more specific primers should be designed in the future when studying the function and mechanism of a candidate eccDNA (*Table 2*).

Based on above results, all the five candidate eccDNAs and the four unexpected eccDNAs detected in high-throughput sequencing were validated by PCR and Sanger sequencing, demonstrating that the presence and expression of these candidate eccDNAs in the PCR reaction was in agreement with the results from the high-throughput sequencing.

Table 1 A comparison of split read counts from high-throughput sequencing and PCR results

Selected candidate eccDNAs	Samples					
	C1	C2	C3	N1	N2	N3
O1						
High-throughput sequencing	8	235	1,360	0	0	0
PCR results	–	+	+	–	–	–
O2						
High-throughput sequencing	41	98	3	0	0	0
PCR results	+	+	–	–	–	–
O3						
High-throughput sequencing	24	1	62	0	0	0
PCR results	+	–	+	–	–	–
O4						
High-throughput sequencing	13	20	2	0	0	0
PCR results	–	+	–	–	–	–
O5						
High-throughput sequencing	1	484	482	0	0	0
PCR results	–	+	+	–	–	–
S1						
High-throughput sequencing	0	0	0	7	5	6
PCR results	–	–	–	–	–	–
S2						
High-throughput sequencing	0	0	0	10	6	4
PCR results	–	–	–	–	–	–
S3						
High-throughput sequencing	0	0	0	9	2	3
PCR results	–	–	–	–	–	–
S4						
High-throughput sequencing	0	0	0	14	4	1
PCR results	–	–	–	–	–	–
S5						
High-throughput sequencing	0	0	0	11	6	2
PCR results	–	–	–	–	–	–

O1–5 indicates candidate up-regulated eccDNA; S1–5 indicates candidate down-regulated eccDNA in ESCC. C1–3 are ESCC samples; N1–3 are matched normal epithelium samples. In high-throughput sequencing, the split read counts are shown. In PCR results, ‘+’ indicates positive bands and ‘–’ indicates negative bands. PCR, polymerase chain reaction; eccDNA, extrachromosomal circular DNA; ESCC, esophageal squamous cell carcinoma.

Table 2 High-throughput sequencing and PCR results of candidate and unexpected eccDNAs

Primer type	EccDNAs ID	Detection type	Samples					
			C1	C2	C3	N1	N2	N3
O1 primer	O1 (chr3:153320590-153321556)	High-throughput sequencing	8	235	1,360	0	0	0
	chr3:153320515-153321586		6	233	1	0	1	0
	O1 (chr3:153320590-153321556)	PCR results	–	–	+	–	–	–
	chr3:153320515-153321586		–	+	–	–	–	–
O2 primer	O2 (chr14:65484268-65484970)	High-throughput sequencing	41	98	3	0	0	0
	chr14:65484250-65484750		38	0	3	0	0	0
	O2 (chr14:65484268-65484970)	PCR results	–	+	–	–	–	–
	chr14:65484250-65484750		+	–	–	–	–	–
O3 primer	O3 (chr1:201630725-201631049)	High-throughput sequencing	24	1	62	0	0	0
	chr1:201630702-201631065		1	0	61	0	0	0
	O3 (chr1:201630725-201631049)	PCR results	+	–	–	–	–	–
	chr1:201630702-201631065		–	–	+	–	–	–
O5 primer	O5 (chr3:137490587-137491130)	High-throughput sequencing	1	484	482	0	0	0
	chr3:137490571-137491158		1	471	0	0	1	1
	O5 (chr3:137490587-137491130)	PCR results	–	–	+	–	–	–
	chr3:137490571-137491158		–	+	–	–	–	–

O1–5 indicates candidate up-regulated eccDNA in ESCC samples. C1–3 are ESCC samples; N1–3 are matched normal epithelium samples. In high-throughput sequencing, the split read counts are shown. In PCR results, '+' indicates positive bands and '-' indicates negative bands. PCR, polymerase chain reaction; eccDNA, extrachromosomal circular DNA; ESCC, esophageal squamous cell carcinoma.

Motif analysis of the flanking eccDNA junctions and the mechanism of eccDNA formation

To explore the mechanism of eccDNA generation, the nucleotide composition from 10 bp upstream to 10 bp downstream of the start and end position for each eccDNA was analyzed. Interestingly, repeat nucleotide patterns were detected between the position flanking the start site and that flanking the end site (Figure 5D and Table S5), and such nucleotide patterns may lead to the generation of eccDNAs.

Discussion

In this study, the extracted DNA in paired healthy and ESCC samples were carefully treated with FastDigest MssI and exonuclease to remove mitochondrial circular DNA and chromosomal linear DNA, and then rolling circle amplification was performed to increase the yield of eccDNAs before sequencing. Thus, only specific eccDNAs, instead of genomic DNA or mitochondrial

DNA, were subsequently investigated by high-throughput sequencing. High-throughput sequencing demonstrated the presence of eccDNAs in paired ESCC samples. Although Møller *et al.* (6) previously showed that eccDNAs were mainly mapped to gene-rich chromosomes, this current study demonstrated that the eccDNAs could be mapped to any region of the human genome and the frequency of eccDNA formation was not associated with the chromosomes nor the coding genes. Interestingly, the eccDNAs were far less frequently mapped to the Y chromosome than other chromosomes and this may be due to less genetic information or the denser structure of the Y chromosome. Mapping the eccDNAs to various genomic regions demonstrated that they originated mainly from regulatory regions such as the 5'-UTR, 3'-UTR, CpG island, upstream or downstream of genes, and less frequently from exons, introns, Alu or LINE repeat region, and this pattern differs from the distribution pattern reported in other studies (11). To explore the

cause of this specific tendency, motif analysis of the flanking eccDNA junctions were performed to study the mechanism of eccDNA formation. Repeat nucleotide patterns were found between the position flanking the start site and that flanking the end site. Previous studies showed that the structure of the 2–15 bp direct repeats flanking the junction site or the dual direct repeats flanking the start and end position possibly contribute to the formation of eccDNAs through circularization mediated by homologous recombination, microhomology end joining, or nonhomologous end joining (4,11). Our results and these latter reports demonstrated that these mechanisms related to direct repeats may participate in the formation of eccDNAs and warrant further investigation.

In agreement with previous reports, the length distribution of the eccDNAs showed that most were less than 1 kb with three peaks at 360, 555, and 736 bp. However, while significant differences in length distribution of eccDNAs have been detected between lung cancer samples and matched normal tissue, before and after surgery, and in maternal and fetal plasma (6,10), no obvious differences were found between ESCC and matched normal epithelium samples in this study. Additionally, no significant differences were observed in the eccDNAs with regards to chromosome distribution and specific genomic regions. Surprisingly, the expression levels of eccDNAs were obviously varied between ESCC and matched normal epithelium. Kumar *et al.* previously speculated that variations in eccDNA abundance in patients with cancer may be difficult to be observed because rolling circle amplification with random primers would result in a massive loss of information related to the level of a specific eccDNA (10). Nevertheless, in our study using a similar method based on screening criteria, an abundant number of significantly up-regulated and down-regulated candidate eccDNAs were identified in ESCC samples. This can be explained by the fact that most of these candidate eccDNAs were detected exclusively in ESCC samples or matched normal tissue and only a small fraction of them were detected in both, thus there was a significant contrast in expression between ESCC and normal samples. Moreover, the expression of the five up-regulated candidate eccDNAs and the four unexpected up-regulated eccDNAs were successfully validated using routine PCR and Sanger sequencing, verifying the authenticity of our high-throughput sequencing data. Interestingly, the absence of the five down-regulated eccDNAs in the PCR may be due to the relatively low expression as indicated by the low split

read counts in high-throughput sequencing, suggesting that high-throughput sequencing is more efficient than routine PCR for the evaluation of candidate eccDNAs with low expression. Improving the efficiency of rolling circle amplification or the development of other advanced techniques may be required to massively increase the yield of products after removing mitochondrial circular DNA and linear DNA. This will facilitate the use of PCR as a convenient and economical method for studying the function and mechanisms of eccDNAs in the future. It is noteworthy that the validation of the four unexpected up-regulated eccDNAs highlights the need for highly specific primers for the PCR amplification of specific products, especially when there are two or more partially-overlapping eccDNAs in the same sample. In addition, the expression of the candidate eccDNAs at altered levels demonstrated that the deletion or amplification of a specific candidate eccDNA in normal esophageal epithelium may lead to tumorigenesis and progression of ESCC. However, the exact mechanism remains to be elucidated.

GO and KEGG pathway analysis of genes associated with the differentially expressed eccDNAs showed enrichment in cancer-related pathways, MAPK pathway, focal adhesion pathway, Rap1 pathway, GTPase related activity, and cytoskeleton function, all of which have been shown to play key roles in ESCC progression (14–17). Most of the eccDNAs were less than 1 kb, and thus too small to carry oncogenes to participate in cancer induction. Møller *et al.* previously identified 25 transcripts across eccDNA junctions by screening mRNA sequencing and speculated a fraction of eccDNAs could be transcribed (6). In our study, we did not observe any transcripts across eccDNA junctions by screening the whole transcriptome sequencing data in the three paired ESCC samples (data not shown). It is possible that the transcription of eccDNAs across eccDNA junction is not a dominant mechanism in ESCC progression. Recently, Paulsen *et al.* reported that shorter eccDNAs (microDNA) carrying microRNA (miRNA) genes or sequences of exons could generate functional small regulatory RNA, such as miRNA or novel si-like RNA, independent of a canonical promoter sequence to modulate related gene expression *in vitro* and *in vivo* (18). In our results, the candidate differentially expressed eccDNAs were mainly derived from the 5'-, 3'-UTRs, and the CpG region, rather than exons, introns or repeat sequences. Previous studies demonstrated that the 5'-, 3'-UTR, and CpG region play important roles in the regulation of RNA translatability and stability, as well as RNA transcription (19–21). Hence,

it is possible that miRNA or si-like RNA generated by the eccDNAs could transcriptionally or posttranscriptionally affect the levels of the corresponding RNA, especially those of oncogenes or tumor suppressor genes, and participate in ESCC induction and progression. However, whether this hypothesis is the dominant mechanism of eccDNA in ESCC remains to be confirmed.

Based on above altered expression of eccDNAs in matched ESCC samples and the discussion, we suggest the deletion or amplification of the eccDNAs in normal esophageal epithelium may function in ESCC initiation by promoting oncogene amplification or generating functional small regulatory RNA *et al.* In the future, we will focus on experimental validation of some representative eccDNAs *in vitro* and *in vivo* to further test our hypothesis in the function of eccDNAs in ESCC. If it can be validated, we think the eccDNAs will have wide applications in the future. For examples, the differences of the extent and nature of eccDNAs between normal and cancer cells may be suggested as biomarker in cancer research. Supposing to be released as extracellular free DNAs, the eccDNAs can be potentially regarded as a basis for detection or monitoring of drug resistance /survival. The functional eccDNAs can be also considered as novel effective therapeutic targets for tumor therapy.

Conclusions

This investigation demonstrated the genome-wide presence of eccDNAs in ESCC samples. Furthermore, the distribution and differential expression of eccDNAs and the potential mechanisms of action were examined in paired ESCC samples. Larger sample sizes are required to further verify these findings. In addition, the specific proteins or regulatory mechanisms involved in the formation of eccDNAs remain to be investigated. Future studies should examine the exact mechanisms leading to the differential expression of eccDNAs in ESCC and the role they play in tumor heterogeneity and patient prognosis. Further investigations should focus on the application of eccDNAs as potential biomarkers and therapeutic targets in patients with ESCC.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 81802397 and 81672292), the Natural Science Foundation of

Shandong Province, China (Grant No. ZR2017BH035), the China Postdoctoral Science Foundation (Grant No. 2020M672073), the Jinan Science and Technology Bureau (Grant No. 2019GXRC051), the Taishan Scholar Program of Shandong Province (Grant No. ts201712087), and the Key Research and Development Program of Shandong Province (Grant No. 2019GSF108072).

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at <https://dx.doi.org/10.21037/atm-21-4372>

Data Sharing Statement: Available at <https://dx.doi.org/10.21037/atm-21-4372>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/atm-21-4372>). The authors report that this work was supported by the National Natural Science Foundation of China (Grant No. 81802397 and 81672292), the Natural Science Foundation of Shandong Province, China (Grant No. ZR2017BH035), the China Postdoctoral Science Foundation (Grant No. 2020M672073), the Jinan Science and Technology Bureau (Grant No. 2019GXRC051), the Taishan Scholar Program of Shandong Province (Grant No. ts201712087), and the Key Research and Development Program of Shandong Province (Grant No. 2019GSF108072). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was approved by the Ethics Committee at Qilu Hospital, Cheeloo College of Medicine, Shandong University, China [reference number KYLL-2021(KS)-013], in accordance with the ethical guidelines of the Declaration of Helsinki (as revised in 2013). Written informed consent was obtained from all three enrolled patients with ESCC undergoing surgery.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-

commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Bray F, Ferlay J, Soerjomataram I, 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.
- Smyth EC, Lagergren J, Fitzgerald RC, et al. Oesophageal cancer. *Nat Rev Dis Primers* 2017;3:17048.
- Yang H, Liu H, Chen Y, et al. Neoadjuvant chemoradiotherapy followed by surgery versus surgery alone for locally advanced squamous cell carcinoma of the esophagus (NEOCRTEC5010): a phase III multicenter, randomized, open-label clinical trial. *J Clin Oncol* 2018;36:2796-803.
- Shibata Y, Kumar P, Layer R, et al. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* 2012;336:82-6.
- Cox D, Yuncken C, Spriggs AI. Minute chromatin bodies in malignant tumours of childhood. *Lancet* 1965;1:55-8.
- Møller HD, Mohiyuddin M, Prada-Luengo I, et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun* 2018;9:1069.
- Tandon I, Pal R, Pal JK, et al. Extrachromosomal circular DNAs: an extra piece of evidence to depict tumor heterogeneity. *Future Sci OA* 2019;5:FSO390.
- Vogt N, Lefèvre SH, Apiou F, et al. Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas. *Proc Natl Acad Sci U S A* 2004;101:11368-73.
- Storlazzi CT, Lonoce A, Guastadisegni MC, et al. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. *Genome Res* 2010;20:1198-206.
- Kumar P, Dillon LW, Shibata Y, et al. Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. *Mol Cancer Res* 2017;15:1197-205.
- Sin STK, Jiang P, Deng J, et al. Identification and characterization of extrachromosomal circular DNA in maternal plasma. *Proc Natl Acad Sci U S A* 2020;117:1658-65.
- Dillon LW, Kumar P, Shibata Y, et al. Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep* 2015;11:1749-59.
- Kumar P, Kiran S, Saha S, et al. ATAC-seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. *Sci Adv* 2020;6:eaba2489.
- Chen J, Zhang W, Wang Y, et al. The diacylglycerol kinase α (DGK α)/Akt/NF- κ B feedforward loop promotes esophageal squamous cell carcinoma (ESCC) progression via FAK-dependent and FAK-independent manner. *Oncogene* 2019;38:2533-50.
- Wang K, Li J, Guo H, et al. MiR-196a binding-site SNP regulates RAP1A expression contributing to esophageal squamous cell carcinoma risk and metastasis. *Carcinogenesis* 2012;33:2147-54.
- Shaverdashvili K, Padlo J, Weinblatt D, et al. KLF4 activates NF κ B signaling and esophageal epithelial inflammation via the Rho-related GTP-binding protein RHO. *PLoS One* 2019;14:e0215746.
- Kajiwara C, Fumoto K, Kimura H, et al. p63-dependent Dickkopf3 Expression promotes esophageal cancer cell proliferation via CKAP4. *Cancer Res* 2018;78:6107-20.
- Paulsen T, Shibata Y, Kumar P, et al. Small extrachromosomal circular DNAs, microDNA, produce short regulatory RNAs that suppress gene expression independent of canonical promoters. *Nucleic Acids Res* 2019;47:4586-96.
- Jia L, Mao Y, Ji Q, et al. Decoding mRNA translatability and stability from the 5' UTR. *Nat Struct Mol Biol* 2020;27:814-21.
- Lee WH, Han MW, Kim SH, et al. Tristetraprolin posttranscriptionally downregulates TRAIL death receptors. *Cells* 2020;9:1851.
- Wu W, Bhagat TD, Yang X, et al. Hypomethylation of noncoding DNA regions and overexpression of the long noncoding RNA, AFAP1-AS1, in Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterology* 2013;144:956-66.e4.

(English Language Editor: J. Teoh)

Cite this article as: Sun Z, Ji N, Zhao R, Liang J, Jiang J, Tian H. Extrachromosomal circular DNAs are common and functional in esophageal squamous cell carcinoma. *Ann Transl Med* 2021;9(18):1464. doi: 10.21037/atm-21-4372

Table S1 The primers of candidate eccDNAs

Candidate eccDNAs	Forward primer	Reverse primer	Product size
O1 (chr3:153320590-153321556)	5'-ACTCTGGCTGGTAGGACCAA-3'	5'-GATTGGGCAAAGAGGTGGGA-3'	256 bp
O2 (chr14:65484268-65484970)	5'-GGGCCTTTCATTGGAAGCAATC-3'	5'-TGACATTGCCCCAGAAGCG-3'	679 bp
O3 (chr1:201630725-201631049)	5'-GAGTGCTGCATCTAGCGGAG-3'	5'-TGGGAGCTGGGGAGTCTAAT-3'	153 bp
O4 (chr5:167233287-167234692)	5'-AAAAGCCTGGCCTCCATTGT-3'	5'-GCTGGGGGCCATCTTCATTA-3'	677 bp
O5 (chr3:137490587-137491130)	5'-GTAAGCAGGATCCACCCGAG-3'	5'-CTCGCGTCTATGCGGGTATT-3'	278 bp
S1 (chr8:38223231-38223598)	5'-CGTTCACCTGTTCAGAGGCTT-3'	5'-CCAGCCTCTCTATATGACTGATTCC-3'	320 bp
S2 (chr1:68141078-68144108)	5'-TGGCCACATCTAAGACCAG-3'	5'-GGTCCGTCTTACTGCTGTG-3'	300 bp
S3 (chr10:70277601-70277938)	5'-ACGGAGAGCTATAATTGTGCGG-3'	5'-CCTCCAGCCTTAGTCTCTGAGT-3'	312 bp
S4 (chr14:31640553-31641482)	5'-AGTTCCAAAAGCAGAGGCC-3'	5'-CCATGCCATCTCCATCCACT-3'	657 bp
S5 (chr13:41890332-41891079)	5'-GCAAGATGATTCTGTCGAACCCA-3'	5'-ATGGAGCAACTTCTCACAGCATC-3'	659 bp

Ox or Sx indicates the up-regulated or down-regulated candidate eccDNAs in ESCC relative to matched normal epithelium. In the form of ChrX: Y-Z, X indicates chromosome id, Y indicates start site and Z indicates end site. EccDNA, extrachromosomal circular DNA.

Table S2 The summary of reads and eccDNA count in each sample

Reads or count	C1	C2	C3	N1	N2	N3
Raw reads	141,510,152	138,644,094	148,054,446	136,927,808	137,162,048	142,572,110
Clean reads	141,495,966	138,630,264	148,034,092	136,914,164	137,144,340	142,555,544
EccDNA count	91,030	103,663	40,755	52,896	71,261	60,837

Cx indicated ESCC, while Nx indicated matched normal epithelium. EccDNA, extrachromosomal circular DNA; ESCC, esophageal squamous cell carcinoma.

Table S3 The genes giving rise to more than 100 eccDNAs

Gene	Number of eccDNAs generated by the gene
<i>LSAMP</i>	192
<i>CSMD1</i>	159
<i>MACROD2</i>	149
<i>MAGI2</i>	142
<i>CNTNAP2</i>	138
<i>RBFOX1</i>	129
<i>WWOX</i>	125
<i>NRXN3</i>	121
<i>LPP</i>	116
<i>NAALADL2</i>	114
<i>DAB1</i>	108
<i>AUTS2</i>	106
<i>LRP1B</i>	106
<i>SHANK2</i>	104
<i>PTPRD</i>	103
<i>IMMP2L</i>	101

EccDNA, extrachromosomal circular DNA.

Table S4 The genes generating the most candidate eccDNAs at differential level (top 10)

Up-regulated eccDNAs in ESCC		Down-regulated eccDNAs in ESCC	
Gene	Count of eccDNAs generated by the gene	Gene	Count of eccDNAs generated by the gene
<i>AUTS2</i>	13	<i>LSAMP</i>	7
<i>SHANK2</i>	11	<i>CSMD1</i>	7
<i>MACROD2</i>	11	<i>BICD1</i>	7
<i>ASIC2</i>	10	<i>KALRN</i>	6
<i>LPP</i>	10	<i>ROBO2</i>	6
<i>EYS</i>	10	<i>FHOD3</i>	5
<i>MAGI2</i>	9	<i>GRID2</i>	5
<i>MYO3B</i>	9	<i>DOCK3</i>	5
<i>TENM4</i>	9	<i>MACROD2</i>	5
<i>DMD</i>	9	<i>RUNX1</i>	5

EccDNA, extrachromosomal circular DNA; ESCC, esophageal squamous cell carcinoma.

Table S5 The nucleotide pattern flanking the eccDNAs junctional sites

Location	Start position				End position			
	A	C	G	T	A	C	G	T
-10	0.297100624739	0.195207984525	0.215521492005	0.29215906197	0.293638279773	0.201812990025	0.214827939336	0.289720790867
-9	0.292950145484	0.193176091939	0.2083150463	0.305547879517	0.293389034282	0.197109835986	0.217959763108	0.291541366624
-8	0.287894796729	0.18256148507	0.225063259589	0.304469621851	0.2895855331361	0.191631853574	0.210655786559	0.308127028506
-7	0.265549396663	0.188521703322	0.210287336704	0.335630276551	0.285375249923	0.182160524933	0.223860379178	0.308603845966
-6	0.317939715102	0.16721121388	0.227078897035	0.287759337224	0.276971342187	0.177397768711	0.220717718645	0.324913170457
-5	0.376680375169	0.157230557497	0.216854413542	0.249223817032	0.332997393759	0.15799996749	0.242380402802	0.266622235949
-4	0.390210070601	0.140444415546	0.230172792146	0.239161884946	0.378062062127	0.148387760963	0.224304686357	0.249245490553
-3	0.301646645752	0.160237758524	0.271016542315	0.267088216649	0.358810557172	0.157219720737	0.237866892071	0.24610283002
-2	0.240586918946	0.217271628819	0.239313599593	0.302817015881	0.267207421014	0.203926158314	0.257047958083	0.271818462589
-1	0.274988215023	0.217271628819	0.246140758682	0.25012868653	0.235764560542	0.240966205563	0.233689320914	0.289579912981
0	0.282156732067	0.264162291325	0.226845906685	0.226835069924	0.287981490813	0.234567098512	0.226119843734	0.2513331566941
1	0.247706670568	0.250513391527	0.161890364494	0.339889573411	0.279913522652	0.262574705917	0.17736525843	0.280141094621
2	0.239443640718	0.219401052249	0.146632205768	0.394523101264	0.248281018872	0.241096246688	0.145597295145	0.365020020915
3	0.24278678132	0.24751160888	0.151698391283	0.358003218518	0.2506655106173	0.214844194477	0.152354015291	0.382131265679
4	0.320204598037	0.226146935635	0.169188922663	0.284459543664	0.273064690042	0.225047004448	0.161034260418	0.340848626712
5	0.316303364272	0.226699610418	0.181342349518	0.275654675791	0.325314130594	0.210991726133	0.183628905975	0.280059818918
6	0.30845213132	0.212774373229	0.191252566958	0.287520928494	0.313252816203	0.218274029162	0.183141251754	0.285326484501
7	0.293638279773	0.219265592744	0.197380754997	0.289715372487	0.306658647464	0.208510107988	0.19013096225	0.294694863917
8	0.287856868068	0.217683425717	0.202262715584	0.292196990632	0.294120515613	0.211750299366	0.196806406693	0.297317359948
9	0.284107348949	0.214383632157	0.206163949349	0.295345069545	0.292982655765	0.212481780696	0.19673596775	0.297794177409
10	0.282693151709	0.213148241465	0.212714771046	0.29144383578	0.294911599127	0.209929723608	0.199217585895	0.29593567299

The left column indicates the position of upstream (negative number) or downstream (positive number) of start/end site. The number in the main text indicates the percentage of one nucleotide at certain position. EccDNA, extrachromosomal circular DNA.