Received: 29 December 2021 | Revised: 18 March 2022 | Accepted: 21 March 2022

DOI: 10.1002/jcla.24425

RESEARCH ARTICLE

WILEY

Differential expression and analysis of extrachromosomal circular DNAs as serum biomarkers in lung adenocarcinoma

Gang Xu[1] | Wenjing Shi[1] | Liqun Ling[1] | Changhong Li[1] | Fanggui Shao[1] | Jie Chen[2] | Yumin Wang[1]

[1]Department of Laboratory Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[2]Department of ICU, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

Correspondence
Jie Chen, 2 Fu Xu Xiang, Department of ICU, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China.
Email: chenjie991300@163.com

Yumin Wang, 2 Fu Xu Xiang, Department of Laboratory Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China.
Email: wangyumin0577@wmu.edu.cn

Funding information
This study was financially supported by the Zhejiang Provincial Health Planning Commission (2021KY784), and the Zhejiang Provincial Research Center for Cancer Intelligent Diagnosis and Molecular Technology (JBZX-202003)

Abstract

Background: Extrachromosomal circular DNAs (eccDNAs) increase the number of proto-oncogenes by enhancing oncogene expression to promote tumorigenesis. However, there are limited reports on differential eccDNA expression and analysis in lung cancer, especially in lung adenocarcinoma (LAD).

Methods: Three LAD and three corresponding NT tissues samples were used for eccDNA next-generation sequencing analysis, and an additional 20 were used for quantitative PCR (qPCR) evaluations. We further performed qPCR amplification using serum samples from LAD patients and healthy medical examiners.

Results: eccDNAs from LAD samples were mainly 200–1000 bp in length. Gene annotation analysis revealed that most eccDNAs were derived from chromosomes 1 and 2. The top-ten increased and top-ten decreased eccDNAs in LAD tissues were CircD-ARPC1B, CircD-ARPC1A, CircD-FAM49B, CircD-SDK1, CircD-KCNG1, CircD-POLR2F, CircD-SS18L1, CircD-SLC16A3, CircD-CSNK1D, CircD-KCTD1, and CircD-TMIGD2, CircD-PDIA5, CircD-VAV2, CircD-GATAD2A, CircD-CAB39L, CircD-KHDC1, CircD-FOXN3, CircD-SULT2B1, CircD-DPP9, and CircD-CSNK1D. qPCR demonstrated that the expression of CircD-DZRN3 was higher in LAD tissues than in normal lung tissues, whereas CircD-LGR6 and CircD-UMODL1 expression levels were lower in LAD than in normal lung tissues. Furthermore, the serum CircD-PDZRN3 level increased, while CircD-LGR6 decreased in LAD. Receiver operating characteristic (ROC) analysis showed that area under curve (AUC) of serum CircD-PDZRN3 (0.991), CircD-LGR6 (0.916) was higher than that of serum carcinoembryonic antigen (CEA) (0.825), CY211 (cytokeratin 19 fragment) (0.842), SCCA(squamous cell carcinoma antigen) (0.857) for the diagnosis of LAD.

Conclusions: Our study first showed that several eccDNAs were aberrantly expressed in LAD, among which CircD-PDZRN3 and CircD-LGR6 clearly distinguished LAD patients from healthy controls, indicating their potential as biomarkers.

KEYWORDS
biomarkers, differential expression, eccDNAs, lung adenocarcinoma

Correction added on 29th April 2022, after first online publication: Figure 1 has been updated.

1 | INTRODUCTION

Lung cancer is the leading cause of cancer deaths worldwide, with the highest incidence and mortality rates among the different types of cancer. Lung adenocarcinoma (LAD) is the most common type of non-small-cell lung cancer (NSCLC), accounting for 40% of lung cancer cases. Despite advances in chemotherapy, radiotherapy, and surgery,[1] the average 5-year survival rate of lung cancer is below 15%.[2-5] Recently, the increasing proportion of LAD patients has drawn great attention in terms of socioeconomic environments. The incidence of LAD has increased significantly in recent years, with a high prevalence in women and non-smoking patients.[6] However, the mechanisms of LAD remain poorly understood.

A class of specialized circular DNA molecules named extrachromosomal circular DNAs (eccDNAs) are widely found in eukaryotic organisms and are independent of the genome or chromosomal genome. They participate in physiological or pathological processes in a specific way. As early as 1965, eccDNA molecules were first found in mammals and higher plants.[7,8] Several eccDNAs are present in human tumors and often carry proto-oncogenes.[9] It was previously concluded that 6 eccDNA molecules exist in healthy human muscle and blood cells.[10] EccDNAs increase the number of proto-oncogenes by enhancing oncogene expression to promote tumorigenesis.[11,12] EccDNAs carry intact genes, especially oncogenic driver genes often present in tumors, and uncontrolled expression of these genes ultimately leads to the malignant growth of tumors. Earlier studies found that oncogene expression was primarily due to an increase in gene copy number.[13] EccDNAs have been found in healthy human somatic cell tissues and shown to be associated with oncogenic gene amplification.[9] Furthermore, they have been reported to serve as a mechanism for the development and intergenerational transmission of herbicide tolerance in weeds.[14] Recent articles found highly open chromatin in eccDNAs and the presence of enhancer sequences, and these features increased their transcriptional activity. As a result, the overexpression of eccDNA on cancer genes contributed to the copy number increase and high transcriptional activity of eccDNA.[10]

eccDNAs have been shown to be prevalent in human tissues and plasma,[15] and it was proposed that because eccDNAs are more stable than linear DNA, eccDNAs circulating in the blood are ideal biomarkers. Therefore, cell-free eccDNAs are not only functionally important but may also be used as potential biomarkers for disease risk assessment, early detection, and outcome prediction. However, there are no literature reports on differential eccDNA expression and analysis in lung cancer, especially in LAD.

2 | MATERIALS AND METHODS

2.1 | Patient samples

From April 2021 to August 2021, LAD samples and corresponding NT samples were prospectively collected from 23 patients in the First Affiliated Hospital of Wenzhou Medical University, China. Three LAD (FAI8, FAI15, FAI22) and three corresponding NT samples (AIP8, AIP15, AIP22) were used for eccDNA next-generation sequencing analysis, and an additional 20 were used for quantitative PCR (qPCR) evaluations. Immediately after excision, the LAD and matched NT samples were quickly frozen in liquid nitrogen. Moreover, we collected 23 LAD and 13 normal control serum samples. Serum samples simultaneously detect lung cancer-related tumor markers, including carcinoembryonic antigen (CEA), CY211 (cytokeratin 19 fragment), squamous cell carcinoma antigen (SCCA). The 23 LAD cases included 12 males and 11 females, ranging from 34 to 80 years old. Inclusion criterion: newly diagnosed patients with untreated lung adenocarcinoma. Exclusion criterion: patients with lung adenocarcinoma who underwent surgery, radiotherapy and chemotherapy, and targeted therapy. The 13 normal control cases included 7 males and 6 females, ranging from 33 to 79 years old. Histopathological results confirmed the diagnosis of LAD. The study was approved by the institutional ethics review committee of the First Affiliated Hospital of Wenzhou Medical University (YS2021-272), and all patients provided written informed consent.

2.2 | Library preparation and sequencing

Circle-Seq eccDNA sequencing services were provided by CloudSeq Biotech Inc., following published procedures with minor modifications.[10] Tissue samples were suspended in L1 solution (Plasmid Mini AX; A&A Biotechnology) supplemented with proteinase K (Thermo Fisher) and then incubated overnight at 50°C with agitation. After lysis, samples were subjected to alkaline treatment, proteins were precipitated, and chromosomal DNA and circular DNA were separated by ion-exchange membrane columns (Plasmid Mini AX; A&A Biotechnology). The column-purified DNA was treated with FastDigest MssI (Thermo Scientific), and mitochondrial circular DNA was removed and incubated at 37°C for 16 h. The remaining linear DNA was removed with exonuclease (Plasmid-Safe ATP-dependent DNase, Epicentre) in a 37°C heating block, and the enzyme reaction was performed continuously for 1 week in accordance with the manufacturer's protocol (Plasmid-Safe ATP-dependent DNase, Epicentre). Additional ATP and DNase were added every 24 h (30 units per day). The eccDNA-enriched sample was used as a template for the phi29 polymerase amplification reaction (REPLI-g Midi Kit) to amplify eccDNA at 30°C for 2 days (46–48 h). Phi29-amplified DNA was sheared by sonication (Bioruptor), and fragmented DNA was amplified with an Illumina (New England Biolabs) NEBNext® Ultra II DNA Library Prep Kit for library preparation. Sequencing of 150 bp paired-end patterns was performed on an Illumina NovaSeq 6000 in accordance with the manufacturer's instructions.

2.3 | Data analysis

Paired-end reads were obtained from the Illumina NovaSeq 6000 sequencer and quality controlled by Q30. After trimming and

removing low-quality reads from 3′ adapters using cutadapt software (v1.9.1), high-quality clean reads were aligned to the reference genome (UCSC hg19) using bwa software v (v0.7.12). Then, eccDNA within all samples was detected using Circle-map software (v1.1.4), and raw soft cut reads of breakpoints were obtained using samtools (v0.2) software. The differentially expressed eccDNAs were then normalized using edgeR (v0.6.9) software and filtered by the *p*-value and fold change, and eccDNAs were annotated using bedtools (v2.27.1) software. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed based on differentially expressed eccDNA-related genes. IGV (v2.4.10) software was used for the visualization of eccDNAs.

2.4 | QPCR verification of differentially expressed eccDNAs

The eccDNA of tissues was extracted and rolling-circle amplified in the same way as the library was prepared and sequenced. For serum plasma samples, a QIAamp Circulating Nucleic Acid Kit was used to extract plasma/serum DNA. The plasma/serum DNA was digested with Plasmid-Safe ATP-dependent DNase at 37°C for 5 min to remove linear DNA. The digested DNA was purified with a MinElute Reaction Cleanup Kit (Qiagen). Five candidate eccDNA genes were selected based on a variety of tissue conditions, including differential variation and gene location, and primers (Table 1) were designed for PCR validation of these genes. The total PCR reaction volume was 20 μl, including 10 μl SYBR Premix (2×), 2 μl DNA template, 1 μl PCR forward primer (10 mM), 1 μl PCR Reverse primer (10 mM), and 6 μl double vaporized water. The initial denaturation step of the quantitative real-time PCR reaction was 95°C for 10 min, followed by 40 cycles of 95°C (5 s) and 60°C (30 s), with a final extension step of 72°C for 5 min. All experiments were performed in triplicate, and all samples were normalized to a pGEX-5X-2 carrier. The median in each triplicate was used to calculate relative eccDNA concentrations (ΔCt = Ct median eccDNA−Ct median pGEX-5X-2 carrier). Expression fold changes were calculated.[16]

2.5 | Statistical methods

All results are expressed as the mean ±standard deviation. The two microarrays were compared using Student's t-test or analysis of variance (ANOVA) for multiple comparisons. In both cases, $p < 0.05$ was considered to indicate a statistically significant difference. Fold change and Student's t-test analyses were performed to understand the statistical significance of the serial results. False discovery rate (FDR) was calculated to correct the *p*-value. We analyzed the area under the ROC curve (AUC) to determine the sensitivity and specificity of the test. The threshold we used to clarify differentially expressed eccDNAs was a fold change of ≥2.0 or ≤0.5 ($p < 0.05$).

3 | RESULTS

3.1 | DNA and DNA library quality control

As shown in Table 2, Table 3, and Figure 1, the quality and concentration of the DNA and DNA library obtained from FAI8, FAI15, FAI22, AIP8, AIP15, and AIP22 samples met the requirements of eccDNA sequencing analysis. After image recognition and base recognition, the original reads were obtained from the Illumina NovaSeq 6000 sequencer. From Table 4, Q30 was used as the QC standard, and Q30 > 80% indicates good sequencing quality.

3.2 | Clean reads, counts, length distribution, and gene annotation of eccDNAs

After image recognition and base recognition, the original reads were obtained from the Illumina NovaSeq 6000 sequencer. The joints and low-quality reads were removed to obtain high-quality clean reads. Next, bwa software was used to compare clean reads to the human reference genome (hg19). The number of clean reads for FAI8, FAI15, FAI22, AIP8, AIP15, and AIP22 was 128,413,440, 135,311,882, 143,348,788, 112,916,090, 122,242,540, and 135,362,092, respectively (Table 5).

The eccDNA counts for the 6 samples were as follows: FAI8 = 18869, FAI15 = 34944, FAI22 = 14516, AIP = 13389, AIP15 = 33309, and AIP22 = 57065 (Figure 2A). These results showed that the distribution of eccDNAs in different LAD tissues was inconsistent, and the distribution of eccDNAs in adjacent and cancerous tissues was also inconsistent.

The length of different eccDNAs from LAD samples ranged from 0 bp to 3000 bp, but most were 200–1000 bp, as shown in

TABLE 1 Primer sequences of genes

Gene	upstream primer (5′–3′)	downstream primer (5′–3′)
CircD-UMODL1	GGCCCCAGCTTCACTTCTAA	TCTGGAGCGACTCATCTCTGG
CircD-SEMA3F	CGCCCGTGCTTCAAATTTTCC	AGGCACCCATGTGGGATTTC
CircD-LGR6	GCGTAGACAAAACTGAGGTCCA	TGCAAAGGGCTAGGCAGATG
CircD-ARPC1A	AGCATTGGCTGAGAGAAGCC	CCCTGAATCCAGAGGTTTGGC
CircD-PDZRN3	GAGGCAGTGTAAGGGCAAAGA	TATCTTCATCAAACAATGTGTGCCT
pGEX−5X−2 carrier	GGGCTGGCAAGCCACGTTTGGTG	CCGGGAGCTGCATGTGTCAGAGG

Figure 2B. Gene annotation analysis revealed that eccDNAs from FAI8, FAI15, FAI22, AIP8, AIP15, and AIP22 were mainly derived from chromosomes 1 and 2, followed by 3–12. The lowest number was from chromosomes X and M, as shown in Figure 2C–2G.

TABLE 2 DNA quantification and quality assurance by Qubit3.0

Sample ID	Sample Name	Conc. (ng/μl)	Volume (μl)	Quantity (μg)
1	FAI8	112	10	1.12
2	FAI15	113.4	10	1.13
3	FAI22	120	10	1.2
4	AIP8	114	10	1.14
5	AIP15	104.6	10	1.05
6	AIP22	103.4	10	1.03

TABLE 3 Sequencing library was determined with an Agilent 2100 Bioanalyzer using the Agilent DNA 1000 chip kit (Agilent, part # 5067–1504)

Sample name	Size (bp)	Conc. (ng/μl)	Conc. (nmol/L)	Volume (μl)[a]	Total amount (ng)
FAI8	304	8.44	42.1	10	84.4
FAI15	336	10.44	47	10	104.4
FAI22	301	11.28	56.8	10	112.8
AIP8	296	8.8	45	10	88
AIP15	298	11.16	56.7	10	111.6
AIP22	298	11.19	57	10	111.9

[a]The libraries were adjusted to 10 nM before cluster generation.

3.3 | Differential expression profile of eccDNAs in LAD

To study the potential biological functions of eccDNAs in LAD, we examined the eccDNA expression profiles in human LAD samples using next-generation sequencing technology. The number of eccDNAs co-expressed in LAD and corresponding NT tissues was 36546 (Figure 3A). The expression profiles of 19,369 eccDNAs indicated that they were differentially expressed (fold change ≥2.0 or ≤0.5; $p < 0.05$) between LAD and normal lung samples, as shown in Figures 3B and 4C. Among these, 7420 eccDNAs were found to be upregulated by more than twofold in the LAD group compared with the normal lung group, whereas 11,949 eccDNAs were downregulated by more than twofold ($p < 0.05$; Table 6, Figures 3B,4C,D). The top-ten increased and top-ten decreased eccDNAs in LAD tissues were CircD-ARPC1B, CircD-ARPC1A, CircD-FAM49B, CircD-SDK1, CircD-KCNG1, CircD-POLR2F, CircD-SS18L1, CircD-SLC16A3, CircD-CSNK1D, CircD-KCTD1 and CircD-TMIGD2, CircD-PDIA5, CircD-VAV2, CircD-GATAD2A, CircD-CAB39L, CircD-KHDC1, CircD-FOXN3, CircD-SULT2B1, CircD-DPP9, and CircD-CSNK1D.

3.4 | GO analysis

mRNA genes associated with the several downregulated eccDNAs were involved in biological processes (including cellular component organization and cellular component organization or biogenesis), cellular components (including cytoplasm and cytoplasmic part), and molecular functions (including protein binding and cytoskeletal

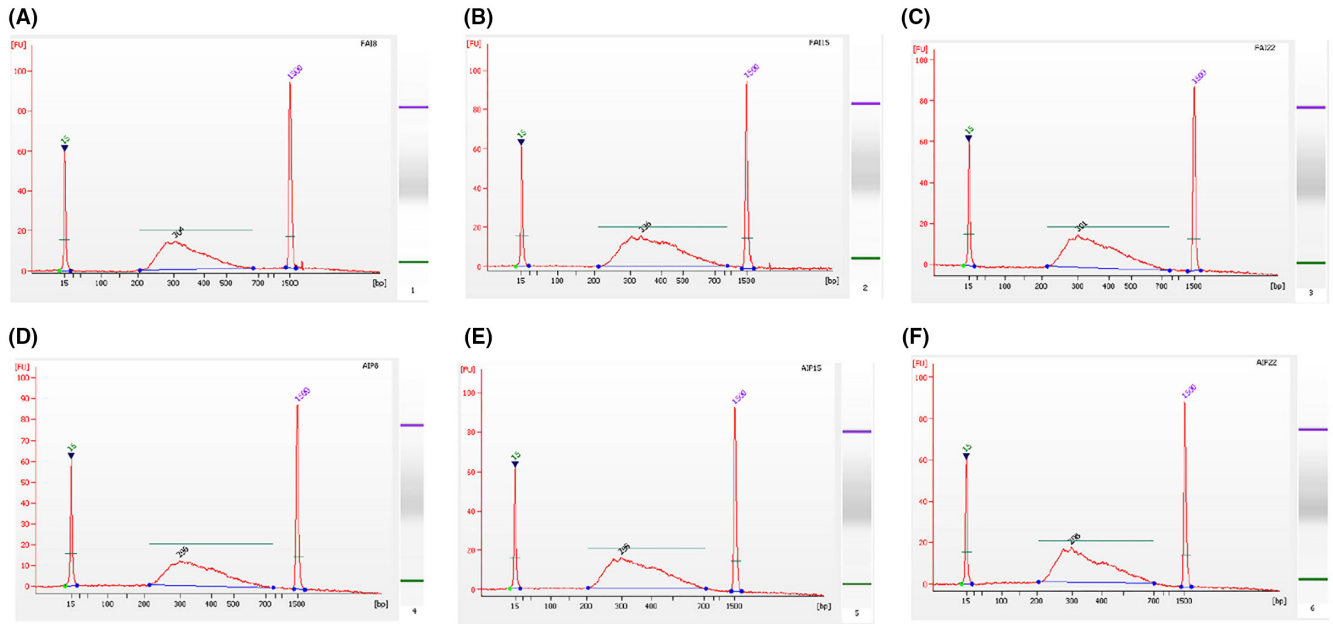

FIGURE 1 The quality and concentration of the DNA and DNA library obtained from FAI8, FAI15, FAI22, AIP8, AIP15, and AIP22 samples met the requirements of eccDNA sequencing analysis. The quality and concentration of the DNA and DNA library obtained from (A) FAI8, (B) FAI15, (C) FAI22, (D) AIP8, (E) AIP15, and (F) AIP22 samples

protein binding), as shown in Figures 4A, 5B,C. The genes corresponding to the several upregulated eccDNAs were involved in biological processes (including cellular component organization and cellular component organization or biogenesis), cellular components (including cytoplasm and DNA packaging complex), and molecular functions (including adenyl nucleotide binding and adenyl ribonucleotide binding), as shown in Figures 4D, 5E,F.

TABLE 4 Q30 of samples

Sample	Q30
FAI8	90.03%
FAI15	89.14%
FAI22	90.34%
AIP15	88.90%
AIP22	88.81%
AIP8	91.08%

TABLE 5 Raw Reads and clean reads of samples

Sample	Raw reads	Clean reads
AIP8	131,263,389	128,413,440
AIP15	147,569,991	135,311,882
AIP22	152,509,390	143,348,788
FAI8	116,259,079	112,916,090
FAI15	127,147,889	122,242,540
FAI22	133,831,858	135,362,092

3.5 | Pathway analysis

Seventy-six upregulated pathways of differentially expressed eccDNA-associated mRNA genes were identified, including pathways related to systemic lupus erythematosus, alcoholism, type I diabetes mellitus, viral myocarditis, viral carcinogenesis, graft-versus-host disease, antigen processing and presentation, arrhythmogenic right ventricular cardiomyopathy (ARVC), allograft rejection, toxoplasmosis, and others. Sixty-eight downregulated pathways of differentially expressed eccDNA-associated mRNAs were identified, including pathways related to systemic lupus erythematosus, type I diabetes mellitus, alcoholism, graft-versus-host disease, cell adhesion molecules (CAMs), ARVC, allograft rejection, axon guidance, viral myocarditis, viral carcinogenesis, and others (Figures 5A, 6B–D). This result suggests that these signaling pathways may be involved in the occurrence and development of LAD.

3.6 | Real-time quantitative PCR validation

According to fold difference, gene locus, nearby encoding gene, and other factors, we initially identified 5 interesting candidate eccDNAs (including CircD-ARPC1A, CircD-PDZRN3, CircD-SEMA3F, CircD-LGR6, and CircD-UMODL1) and verified the expression of these eccDNAs by qPCR. In Figure 6A, the qPCR results show that the CircD-PDZRN3 ($t = 14.932$, $p = 0.0002$) expression levels in LAD were higher than in normal lung tissues, whereas CircD-LGR6 ($t = 2.000$, $p = 0.002$) and CircD-UMODL1 ($t = 1.493$, $p = 0.04$)

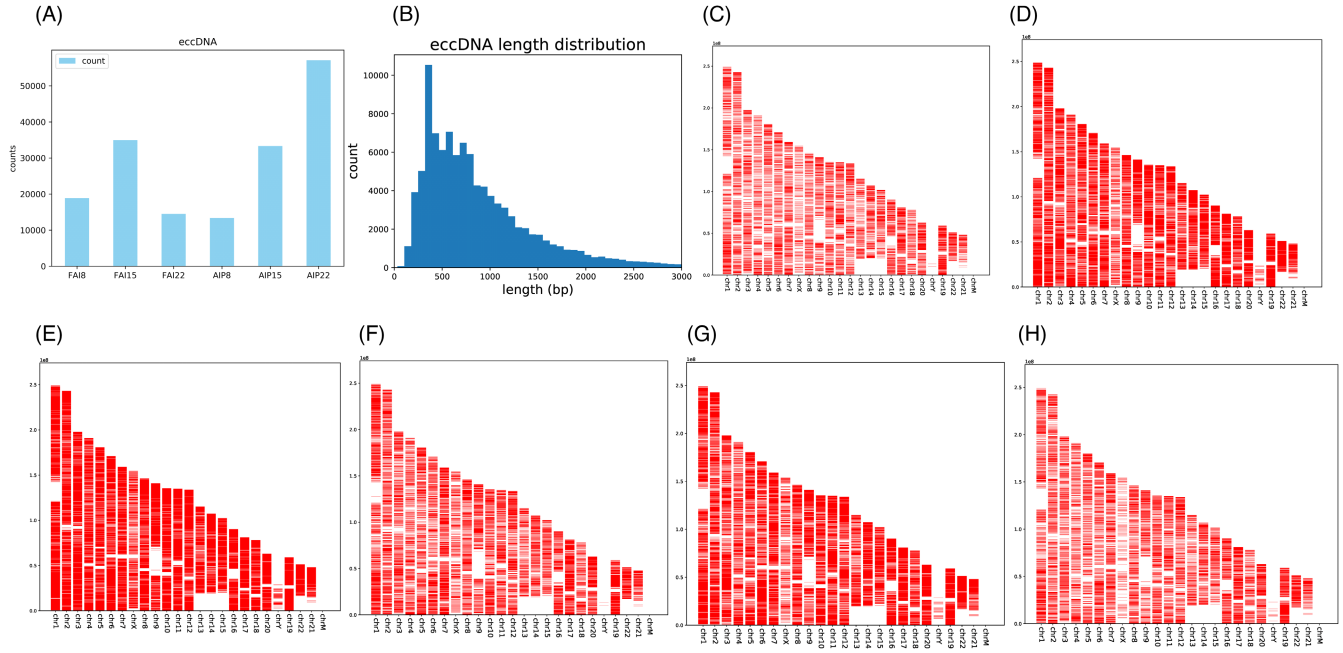

FIGURE 2 The eccDNA counts and lengths in LAD and paracancerous tissues and distribution of differentially expressed eccDNAs on each chromosome. (A) eccDNA counts for the 6 samples were as follows: FAI8 = 18869, FAI15 = 34944, FAI22 = 14516, AIP = 13389, AIP15 = 33309, and AIP22 = 57065. (B) Length distribution of different eccDNAs from LAD samples was from 0 bp to 3000 bp, but most were 200–1000 bp. Chromosome gene annotation of eccDNAs for (C) FAI8, (D) FAI15, (E) FAI22, (F) AIP8, (G) AIP15, and (F) AIP22

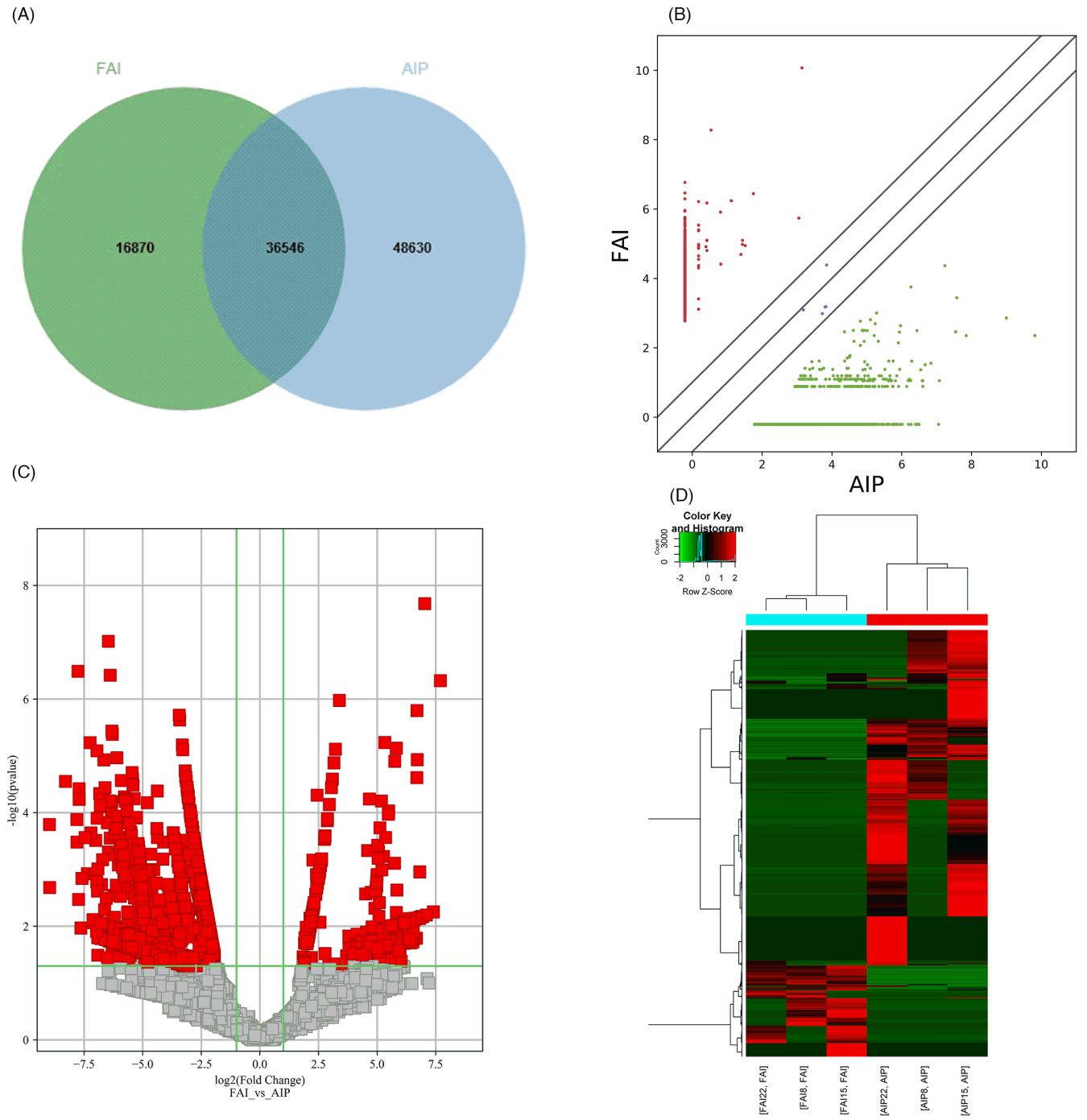

FIGURE 3 EccDNA co-expressed in LAD and paracancerous tissues and scatter plots and volcano plots for differentially expressed eccDNA genes in LAD and paracancerous tissues. (A) The number of eccDNA co-expressed in LAD and paracancerous tissues was 36546. (B) The values of the X and Y axes in the scatter plot are averaged normalized values in each group (log2-scaled). The eccDNAs above the top green line and below the bottom green line are those with a >2-fold change in expression between tissues. (C) Scatter plots were plotted with the expression values of both samples (or the mean expression values of both samples). Red dots indicate up-adjustment, green dots indicate down-adjustment, and the default multiplicative change threshold was 2.0. Volcano plots were plotted using the fold change and P-value when comparing two samples. Red rectangles represent differentially expressed eccDNAs. (D) Heat map and hierarchical clustering of eccDNAs

expression levels in LAD were lower than in normal lung tissues. In contrast, there were no statistically significant differences in CircD-ARPC1A (t = 0.194, $p = 0.867$) and CircD-CircD-SEMA3F (t = 1.503, $p = 0.272$) expression levels between LAD and normal lung tissues. Furthermore, we found that the serum CircD-PDZRN3 (t = 6.845, $p = 0.000$, Figure 6B) levels in LAD were higher than in the normal healthy group, and 100% (23/23) of LAD patients showed increased levels. In contrast, the serum CircD-LGR6 (t = 3.248, $p = 0.004$, Figure 6C) level in LAD patients was lower than in the normal healthy group, and 95.65% (22/23) of LAD patients showed decreased

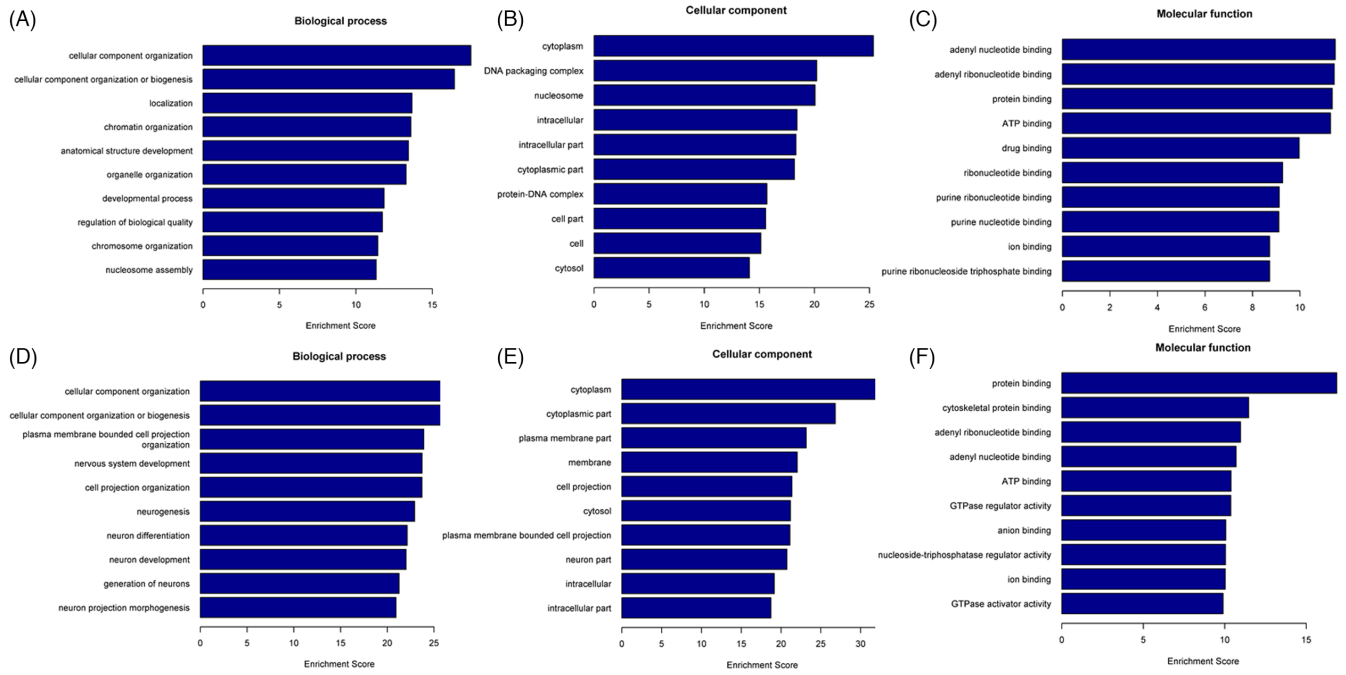

FIGURE 4 GO analysis of eccDNA-associated mRNAs. The genes corresponding to the several downregulated eccDNA-associated mRNAs were involved in (A) biological processes (including cellular component organization and cellular component organization or biogenesis), (B) cellular components (including cytoplasm and cytoplasmic part), and (C) molecular functions (including protein binding and cytoskeletal protein binding). The genes corresponding to the several upregulated eccDNAs were involved in (D) biological processes (including cellular component organization and cellular component organization or biogenesis), (E) cellular components (including cytoplasm and DNA packaging complex), and (F) molecular functions (including adenyl nucleotide binding and adenyl ribonucleotide binding)

TABLE 6 Differential expression profiles of eccDNAs in lung adenocarcinoma

ecc_chr	ecc_start	ecc_end	logFC	*p* value	Regulation	Gene_id	Gene_name
chr4	1.46E+08	1.46E+08	5.990492	0.009429	Up	ENSG00000164162	ANAPC10
chr20	50160632	50161017	5.990492	0.009429	Up	ENSG00000101096	NFATC2
chr12	52908737	52909744	5.990492	0.009429	Up	ENSG00000186081	KRT5
chr10	6060272	6060720	5.990492	0.009429	Up	ENSG00000134460	IL2RA
chr11	70863368	70863785	−6.54754	0.003236	Down	ENSG00000162105	SHANK2
chr19	18070141	18070842	−6.49772	0.035883	Down	ENSG00000105642	KCNN1
chr9	79397115	79398091	−6.47718	0.009942	Down	ENSG00000106772	PRUNE2
chr1	46498882	46499542	−6.46326	9.64E−08	Down	ENSG00000086015	MAST2

Note: The eccDNAs were aberrantly expressed and as potential biomarkers in LAD.

levels. The CircD-UMODL1 (t = 0.243, $p = 0.957$, Figure 6D) expression level in LAD was not significantly different compared with normal lung tissue. Therefore, these eccDNAs clearly distinguished LAD patients from normal healthy controls, indicating their potential as biomarkers in LAD.

3.7 | Serum CircD-PDZRN3 and CircD-LGR6 could serve as potential serum biomarker of LAD

ROC curve analysis showed that serum CircD-PDZRN3 and CircD-LGR6 had excellent diagnostic ability (cut-off = 11.8978, AUC was 0.847, sensitivity = 64.90%, specificity = 95.60%) to distinguish LAD patients between healthy subjects (Figure 6E,6F and Table 7). These results indicated that the combination of serum CircD-PDZRN3 and CircD-LGR6 may serve as a potential serum biomarker of LAD.

4 | DISCUSSION

This study identified 19,369 differentially expressed eccDNAs between LAD and normal lung samples. Among these, 7,420 eccDNAs were found to be upregulated by more than twofold in the LAD group compared with the normal lung group, and 11,949 eccDNAs were downregulated by more than twofold. Therefore, the number

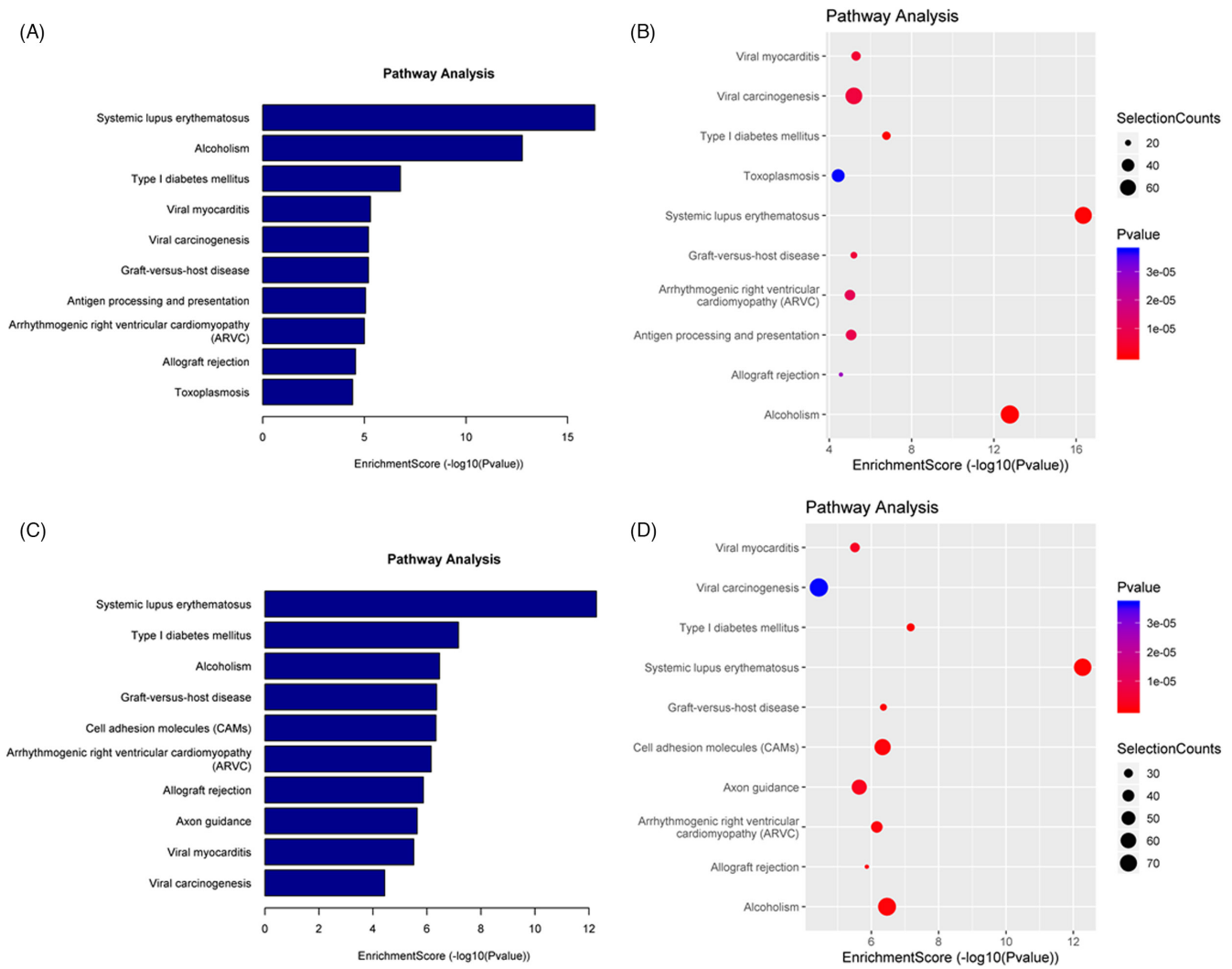

FIGURE 5 Pathway analysis of eccDNA-associated mRNAs. (A)(B) Seventy-six upregulated pathways of differentially expressed eccDNA-associated mRNAs were identified, including pathways related to systemic lupus erythematosus, alcoholism, type I diabetes mellitus, viral myocarditis, viral carcinogenesis, graft-versus-host disease, antigen processing and presentation, arrhythmogenic right ventricular cardiomyopathy (ARVC), allograft rejection, toxoplasmosis, and others. (C)(D) Sixty-eight downregulated pathways of differentially expressed eccDNA-associated mRNAs were identified, including pathways related to systemic lupus erythematosus, type I diabetes mellitus, alcoholism, graft-versus-host disease, cell adhesion molecules (CAMs), arrhythmogenic right ventricular cardiomyopathy (ARVC), allograft rejection, axon guidance, viral myocarditis, viral carcinogenesis, and others

of downregulated eccDNAs was higher than that of upregulated eccDNAs in LAD.

Based on gene annotation analysis, we found that most eccDNAs from FAI8, FAI15, FAI22, AIP8, AIP15, and AIP22 were derived from chromosomes 1 and 2. Of all 22 autosome pairs in the human body, chromosome 1 contains the highest number of genes and is the largest. Approximately 245,522,847 nucleotide base pairs are found in chromosome 1, which contains roughly 8% of the DNA in human cells and 3141 genes. More than 350 diseases are thought to be linked to alterations in genes on chromosome 1, including cancer, Parkinson's disease, Alzheimer's disease, and mental retardation. Chromosome 2 is the second largest chromosome in humans, with 237 million base pairs, and it occupies approximately 8% of the total DNA in the cell. This study showed that most eccDNAs in LAD were derived from the genes on chromosomes 1 and 2. Most eccDNAs of LAD were 200–1000 bp in length. Pathway analysis revealed that the upregulated or downregulated pathways of differentially expressed eccDNA-associated mRNAs were related to systemic lupus erythematosus, alcoholism, type I diabetes mellitus, and others.

We initially identified 5 interesting candidate eccDNAs (including CircD-ARPC1A, CircD-PDZRN3, CircD-SEMA3F, CircD-LGR6, and CircD-UMODL1) and verified their expression by qPCR. The results showed that CircD-PDZRN3 (100%) expression in LAD was higher than in normal lung tissue, whereas CircD-LGR6 and CircD-MODL1 expression levels in LAD were lower than in normal lung tissue. Furthermore, we found that the serum CircD-PDZRN3 levels in LAD were higher than in the normal healthy group, and the serum CircD-LGR6 (95.6%) levels in LAD were lower than in the normal healthy group. Therefore, these eccDNAs clearly distinguished LAD

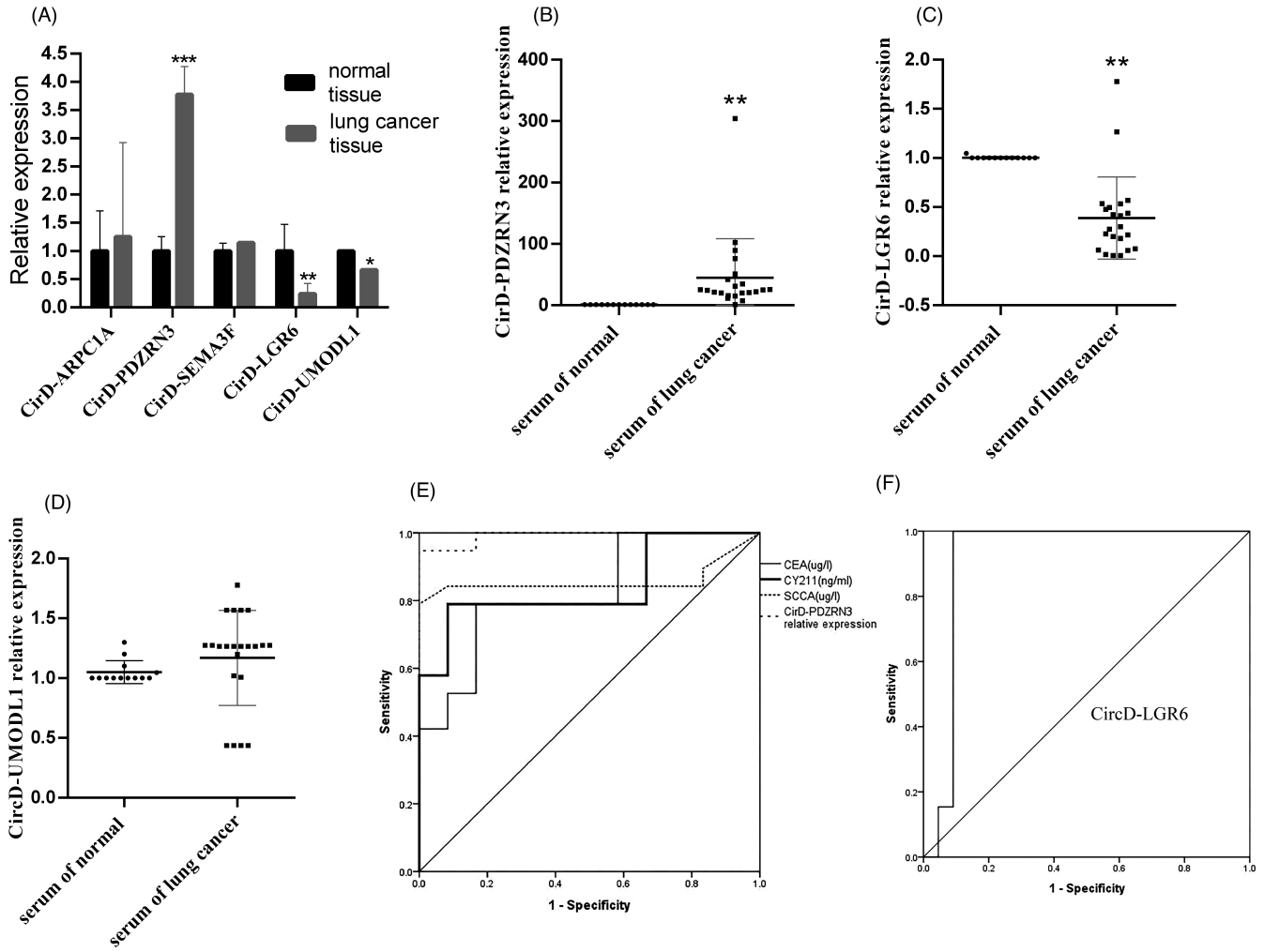

FIGURE 6 Real-time quantitative PCR validation of eccDNA expression levels in LAD and normal lung samples. (A) The qPCR results showed that the CircD-PDZRN3 (t = 14.932, $p = 0.0002$) expression levels in LAD were higher than in normal lung tissue, whereas CircD-LGR6 (t = 2.000, $p = 0.002$) and CircD-UMODL1 (t = 1.493, $p = 0.04$) expression levels in LAD were lower than in normal lung tissue. CircD-ARPC1A (t=0.194, p=0.867) and CircD-CircD-SEMA3F (t=1.503, p=0.272) expression levels in LAD were not significantly different compared with normal lung tissue. (B) Furthermore, the serum CircD-PDZRN3 (t = 6.845, $p = 0.000$, 23/23) levels in LAD were higher than in the normal healthy group. (C) Serum CircD-LGR6 (t = 3.248, $p = 0.004$, 22/23) levels in LAD were lower than in the normal healthy group. (D) CircD-UMODL1 (t = 0.243, $p = 0.957$) expression levels in LAD were not significantly different compared with normal lung tissues. (E) ROC curve analysis showed that serum CircD-PDZRN3 had good diagnostic ability (cut-off = 4.41, AUC was 0.991) to distinguish LAD patients between healthy subjects. (F) ROC curve analysis showed that serum CircD-LGR6 had good diagnostic ability (cut-off = 0.73, AUC = 0.916) to distinguish LAD patients between healthy subjects

TABLE 7 Diagnostic performance of different indicators for lung adenocarcinoma

Parameters	Cut-off	AUC	Specificity	Sensitivity
Serum CEA (ng/ml)	2.75	0.825	0.833	0.789
Serum CY211(ng/ml)	2.35	0.842	0.917	0.789
Serum SCCA (ng/ml)	1.10	0.857	1.000	0.789
Serum CircD-PDZRN3	4.41	0.991	1.000	0.947
Serum CircD-LGR6	0.73	0.916	0.909	1.000

patients from normal healthy controls, thereby providing potential biomarkers in LAD.

In this research, LAD patients had higher levels of serum CircD-PDZRN3 and lower levels of serum CircD-LGR6 than healthy subjects. And the diagnostic ability of serum CircD-PDZRN3 (AUC = 0.991) and serum CircD-LGR6 (AUC = 0.916) for LAD patients was better than serum CEA (AUC = 0. 0.825), CY211 (AUC = 0.842), and SCCA (AUC = 0.857), in LAD patients. The results indicated that serum CircD-PDZRN3 and CircD-LGR6 may become a potential serum biomarker of LAD. However, the specific

mechanisms of eccDNA PDZRN3 and LGR6 in LAD require further in-depth study.

In summary, our study showed for the first time that several eccDNAs were aberrantly expressed in LAD, among which PDZRN3 and LGR6 clearly distinguished LAD patients from normal healthy controls, thereby providing potential biomarkers in LAD.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

All authors have reviewed the data analyses. GX, WS, and YW designed and conducted experiments, analyzed data, and wrote the manuscript. GX, FS, and YW drafted and revised the manuscript. GX, WS, LL, CL, FS, FS and YW designed the study, supervised the research, and revised the manuscript. All authors read and approved the final manuscript. All authors read and approved the final manuscript.

ETHICAL APPROVAL

This study was conducted in accordance with the Declaration of Helsinki. The study was approved by the institutional ethics review committee of the First Affiliated Hospital of Wenzhou Medical University (YS2021-272).

DATA AVAILABILITY STATEMENT

Requests for data, resources, and reagents should be directed to the corresponding author Yumin Wang (wangyumin0577@wmu.edu.cn).

ORCID

Wenjing Shi https://orcid.org/0000-0003-3666-1110
Yumin Wang https://orcid.org/0000-0001-8243-3659

REFERENCES

1. Gridelli C, Rossi A, Maione P. Treatment of non-small-cell lung cancer: state of the art and development of new biologic agents. *Oncogene*. 2003;22(42):6629-6638.
2. Ogawa E, Takenaka K, Katakura H, et al. Perimembrane Aurora-A expression is a significant prognostic factor in correlation with proliferative activity in non-small-cell lung cancer (NSCLC). *Ann Surg Oncol*. 2008;15(2):547-554.
3. Rachet B, Woods LM, Mitry E, et al. Cancer survival in England and Wales at the end of the 20th century. *Br J Cancer*. 2008;99(suppl 1):S2-10.
4. Chen CH, Lai JM, Chou TY, et al. VEGFA upregulates FLJ10540 and modulates migration and invasion of lung cancer via PI3K/AKT pathway. *PLoS One*. 2009;4(4):e5052.
5. Stewart DJ. Tumor and host factors that may limit efficacy of chemotherapy in non-small cell and small cell lung cancer. *Crit Rev Oncol Hematol*. 2010;75(3):173-234.
6. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115-132.
7. Gaubatz JW. Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells. *Mutat Res*. 1990;237(5–6):271-292.
8. Buchowicz J. Nuclear extrachromosomal DNA of higher plants. *Acta Biochim Pol*. 1997;44(1):13-19.
9. Kumar P, Dillon LW, Shibata Y, Jazaeri AA, Jones DR, Dutta A. Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. *Mol Cancer Res*. 2017;15(9):1197-1205.
10. Moller 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(1):1069.
11. Wu S, Turner KM, Nguyen N, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature*. 2019;575(7784):699-703.
12. Morton AR, Dogan-Artun N, Faber ZJ, et al. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell*. 2019;179(6):1330-1341.e13.
13. Kim H, Nguyen NP, Turner K, et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat Genet*. 2020;52(9):891-897.
14. Koo DH, Molin WT, Saski CA, et al. Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed Amaranthus palmeri. *Proc Natl Acad Sci USA*. 2018;115(13):3332-3337.
15. Zhu J, Zhang F, Du M, Zhang P, Fu S, Wang L. Molecular characterization of cell-free eccDNAs in human plasma. *Sci Rep*. 2017;7(1):10968.
16. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402-408.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Xu G, Shi W, Ling L, et al. Differential expression and analysis of extrachromosomal circular DNAs as serum biomarkers in lung adenocarcinoma. *J Clin Lab Anal*. 2022;36:e24425. doi:10.1002/jcla.24425