

# Identification and Characterization of Extrachromosomal Circular DNA in Plasma of Lung Adenocarcinoma Patients

Xiaoqiong Wu<sup>1,\*</sup>, Pu Li<sup>1,\*</sup>, Maimaitiaili Yimiti<sup>1</sup>, Zhiqiu Ye<sup>2</sup>, Xuqian Fang<sup>2</sup>, Peizhan Chen<sup>3</sup>, Zhidong Gu<sup>1,4</sup>

<sup>1</sup>Department of Laboratory Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, People's Republic of China;

<sup>2</sup>Department of Pathology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, People's Republic of China; <sup>3</sup>Clinical Research Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, People's Republic of China; <sup>4</sup>Department of Laboratory Medicine, Ruijin-Hainan Hospital, Shanghai Jiao Tong University School of Medicine (Hainan Boao Research Hospital), Shanghai, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Zhidong Gu, Department of Laboratory Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 201821, People's Republic of China, Tel +86 13801653534, Email guzhidongruijin@163.com; Peizhan Chen, Clinical Research Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 201821, People's Republic of China, Tel +86 13918550745, Email pzchen@me.com

**Background:** Chromosome is the basic framework for eukaryotic cells to store genetic information, but certain genes exist in circulation, such as extrachromosomal circular DNA (eccDNA). The unique genetic characteristics and structure of eccDNA provide a new vision on the early diagnosis of cancer; however, whether eccDNA contributes to the early diagnosis and progression of lung cancer remains unclear.

**Methods:** We performed next-generation sequencing (NGS) analysis of eccDNA from the plasma of 6 lung adenocarcinoma (LUAD) patients. The data of plasma eccDNA of healthy people were obtained from public available database. We compared size distribution, chromosome origin, formation and expression patterns of eccDNA between LUAD patients and those of 6 healthy people and 4 healthy gravidas.

**Results:** A total number of 716,059 eccDNA ranging from 22 bp to 3,297,519 bp were detected with an average size less than 800bp and distinctive bimodality in size around 191 bp and 320 bp. After comparison of eccDNA abundance in each sequencing sample, nine eccDNA were ranked on top with higher frequency in lung adenocarcinoma patients than healthy people. Among them, four eccDNA (DOCK1, PPIC, TBC1D16, and RP11-370A5.1) were uniquely expressed in lung adenocarcinoma patients, which may serve as potential biomarkers for early diagnosis LUAD.

**Conclusion:** Cancer-specific eccDNA was presented in LUAD compared to normal people, which might serve as a promising biomarker in LUAD.

**Keywords:** eccDNA, lung adenocarcinoma, plasma, early diagnosis

## Introduction

Extrachromosomal circular DNA (eccDNA) is a type of cell free DNA (cfDNA) which is more structurally stable than linear cfDNA currently used for cancer-related detections under clinical settings.<sup>1,2</sup> Although eccDNA was first discovered in 1965 by Hotta and Bassel et al in boar sperm using an electron microscope, further research on eccDNA was understudied due to its low abundance in blood.<sup>3-7</sup> With the advance of medical technologies such as NGS and rolling circle amplification, improved sequencing with lower cost has unlocked the possibility of eccDNA to be used in clinical diagnosis for cancer.<sup>8</sup>

Research on the possibility of nucleic acids in body fluids as disease biomarker has been ongoing since the discovery of acellular nucleic acids in human plasma in the 1940s.<sup>9-11</sup> Previous studies have demonstrated structural, functional and

quantitative diversity of eccDNA can effectively drive cancer evolution and reveal details about life stage, tissue type, and disease condition of a person.<sup>9,11</sup> With its stability over linear cfDNA, eccDNA would be a more desirable biomarker for cancer diagnostic approaches.

Modern techniques such as Southern Blot and sequencing have helped researchers in discovering that normal, aged, and damaged cells all contain substantial amounts of eccDNA with vastly different characteristics.<sup>12</sup> eccDNA is homologous with genomic DNA, suggesting that genomic DNA mutations play a role in the formation of eccDNA.<sup>12</sup> In addition, eccDNA have been linked to DNA repair, hyper transcription, homologous recombination and other processes.<sup>13–15</sup> Building upon these findings, we hypothesized that identification of eccDNA in lung cancer patients that differ from healthy individuals will point us to new biomarkers of lung cancer.

This study utilized Next-Generation Sequencing (NGS) to examine eccDNA in plasma. Discrepancies in eccDNA expression profiles of healthy individuals, pregnant women and LUAD patients were compared. To our knowledge, there are few studies that analyze the characteristics of eccDNA in cancer patients and link eccDNA to early diagnosis of lung adenocarcinoma. Our study may be the first study to explore the potential of eccDNA as a LUAD biomarker.

## Materials and Methods

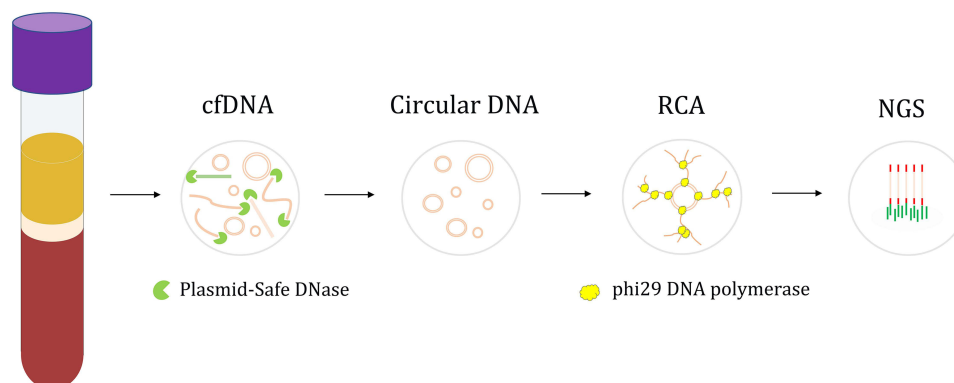
### The Workflow for Extracting and Detecting eccDNA from Plasma

The workflow for extracting and detecting eccDNA from plasma is described in [Figure 1](#). eccDNA from plasma of 6 patients with lung adenocarcinoma were studied to understand the relationship between eccDNA and lung adenocarcinoma. Circle-seq was used to detect the genome after extracting and enriching eccDNA from plasma.<sup>16</sup>

### Patient Recruitment and eccDNA Extraction

In March 2020, primary lung adenocarcinoma patients attending respiratory clinical at Ruijin Hospital were recruited with written informed consent. Six primary LUAD patients without any serious complications and had never been treated were selected, including 3 male subjects and 3 female subjects, aging from 46 years old to 70 years old. Only one male subject had formed the habit of smoking. All peripheral blood samples were collected in fasting state before related treatment using EDTA Vacuum Tubes and processed within 2 hours to ensure freshness. Details of the patients are shown in [Supplementary Table 1](#). Blood samples were first centrifuged at 3000 rpm for 5 min at 4°C. The plasma portion gained was then further centrifuged at 10,000 rpm for 10 min at 4°C to remove residual cells and debris. Thereafter, the plasma samples were stored at –80°C for later use.

cfDNA, which included both linear DNA and circular DNA, was extracted using QIAamp Circulating Nucleic Acid Kit (Catalogue No. #55,114) from a 4 mL plasma sample. After adding 20 µL elution buffer and centrifuging at 14,000 rpm for 1 minute, DNA from the column was eluted. To remove linear DNA, 10 µL eluted DNA was used with an Epicentre plasma-safe ATP-dependent DNase Kit (Biosearch Technologies), then put in an iron bath at 37°C for 1 hour



**Figure 1** Working flow chart of eccDNA identification. cfDNA consisting of linear and circular DNA was isolated from the plasma of 6 lung adenocarcinoma patients. To obtain clean eccDNA, linear cfDNA was digested by exonuclease using Plasmid-Safe ATP-dependent DNase. eccDNA was then sequenced by next generation sequencing (NGS) after rolling circle amplification using phi29 DNA polymerase.

for sufficient digestion, followed by 70°C for 5 minutes to inactivate the DNase. The leftover eccDNA was then cleaned with MinElute Reaction Cleanup Kit (Qiagen) after linear DNA was removed. A 20 µL of pure circular DNA is ultimately yielded, and quality control was accomplished after the purified DNA was quantified using a Qubit 3.0 Fluorometer ([Supplementary Table 2](#)).

## Ethical Statement

All experiments were performed in accordance with relevant guidelines and regulations. Ethical approval for this study was obtained from the ethics committee of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine [Agreement Number: (2020) (013)-1]. This study was performed in accordance with the Declaration of Helsinki.

## Library Preparation and Sequencing

Nextera XT DNA Library Preparation Kit (Illumina) was used to process the enhanced eccDNA. After building DNA libraries, eccDNA quality control was performed ([Supplementary Table 3](#)). DNA libraries were sequenced on an Illumina NovaSeq 6000 using the 150bp paired-end mode. Original data were obtained, and eccDNA libraries were constructed based on images and base recognitions.

## Identification of eccDNA

The quality assessment of NGS result was measured using Q30, which represents the percentage of bases with mass value  $\geq 30$ . The sequencing quality score of a given base, Q, is defined by the following equation:  $Q = -10\log_{10}(e)$ . Higher Q scores indicate a smaller probability of error. Q30 >80% was assumed to be of usable sequencing quality. Before read mapping, the cutadapt software (v1.9.1) was used to find adapters and remove them error-tolerantly. Clean reads were then compared to the human reference genome (hg38) using BWA software (v0.7.12). Statistics data of Q30 quality control is shown in [Supplementary Table 4](#). Python package Circle-Map (v1.1.4) was used to identify eccDNA in all samples using the following four bioinformatics criteria: (i) At least one split reads; (ii) Circle score above 30; (iii) Coverage increase in the start coordinate and end coordinate greater than 0.33; (iv) Coverage continuity less than 0.2.<sup>17</sup> The total number of split reads was taken as the original count of each kind of eccDNA.

## General Characteristics of eccDNA

After locating the eccDNA molecules in the reference genome, we evaluate their general characteristics, including the length distribution and the chromosome origin preference, all plots were generated by the R package ggplot2 (v3.3.5.9000). The standardized chromosome origin (counts/mega bases, counts/Mb) was obtained through dividing the counts of eccDNA by the chromosome size (Mb). R package RIdeogram (v0.2.2) was used to visualize the chromosome origin of eccDNA.

## Genomic Annotation of Plasma eccDNA

Genomic data in R package of TxDb.Hsapiens.UCSC.hg38.knownGene (v3.13.0) was used as the reference, and all plasma eccDNA was annotated using ChIPseeker (v1.28.3). Furthermore, the amount of these eccDNA molecules with the junctional positions mapped to the 9 genomic element classes were obtained, including 5'-UTR, 3'-UTR, distal intergenic, exon, intron, downstream ( $\leq 300$ bp), promoter ( $\leq 1$ kb), promoter (1–2kb) and promoter ( $> 2$ kb). The theoretical distribution of plasma eccDNA in each kind of element was predicted as the percentage of genome covered by its respective genomic elements. Furthermore, to reveal the prevalence of epigenomic elements, including CpG islands, DNase Clusters, H3K4Me1 (GM12878), H3K4Me3 (K562) and H3K27Ac (K562), all marks of these elements were downloaded from UCSC ENCODE Regulation Super-track (hg38) and plasma eccDNA was analyzed using bedtools from Galaxy (<https://usegalaxy.org/>). Data analysis results were visualized by [http://www.ehbio.com/Cloud\\_Platform/front/](http://www.ehbio.com/Cloud_Platform/front/).

## Junctional Nucleotide Motif Patterns of eccDNA

To determine if there was any preference in eccDNA generation, DNA motif patterns flanking the eccDNA were explored. Through BSgenome (v1.60.0), the base compositions from 10 bp upstream to 10 bp downstream of the start

and end positions for each eccDNA loci were scanned. All nucleotide motifs were inferred from the reference genome and visualized using ggplot2 (v3.3.5.9000) and ggseqlogo (v0.1).

## Differentially Expressed eccDNA Between Lung Adenocarcinoma Patients and Healthy People

The sequencing data of plasma eccDNA from healthy people were used as contrast to explore plasma eccDNA molecules conducive to lung adenocarcinoma diagnosis. Through searching eccDNA plasma sequencing datasets in GEO database, we identified two datasets, one is performed in 6 healthy adults,<sup>18</sup> the other is performed in 4 healthy gravidas.<sup>19</sup> The raw sequencing data downloaded from Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>) (SRP115110) and the European Genome-Phenome Archive (EGA) (<https://ega-archive.org/datasets/EGAD00001005286>) (accession No.: EGAS00001003827) were named as group B and group C ([Supplementary Table 1](#)),<sup>18,19</sup> and our inhouse sequencing data of 6 lung adenocarcinoma patients were named as group A in comparisons.

For group B, cfDNA of 4 healthy gravidas was extracted using QIAamp DNA Blood Maxi Kit (Qiagen), and cfDNA was treated with Plasmid-Safe™ ATP-dependent DNase (Epicentre) to eliminate linear DNA. Before purifying DNA using Agencourt AMPure XP (Beckman Coulter, #A63881) and quantifying using iQuant High Sensitivity dsDNA Quantitation Kit (GeneCopoeia, #N011), eccDNA was amplified using the REPLI-g Single Cell Kit (Qiagen). eccDNA was subjected to DNA library preparation using the ThruPLEX DNA-seq kit (Rubicon Genomics) and indexed libraries were sequenced with 125 bp paired-end sequencing (Illumina HiSeq 2500).

For data of group C, cfDNA of 6 healthy people was extracted using QIAamp Circulating Nucleic Acid Kit (Qiagen), and cfDNA was treated with exonuclease V (New England Biolabs) to eliminate linear DNA. After purification using MinElute Reaction Cleanup Kit (Qiagen), the resultant circular DNA was digested with MspI (New England Biolabs). The sequencing libraries were prepared using TruSeq Nano DNA LT Library Prep Kit (Illumina) and DNA libraries were sequenced with  $2 \times 250$ -bp paired-end reads on a HiSeq 2500 platform.

## Analysis of the Differentially Expressed eccDNA and Their Mapping Genes

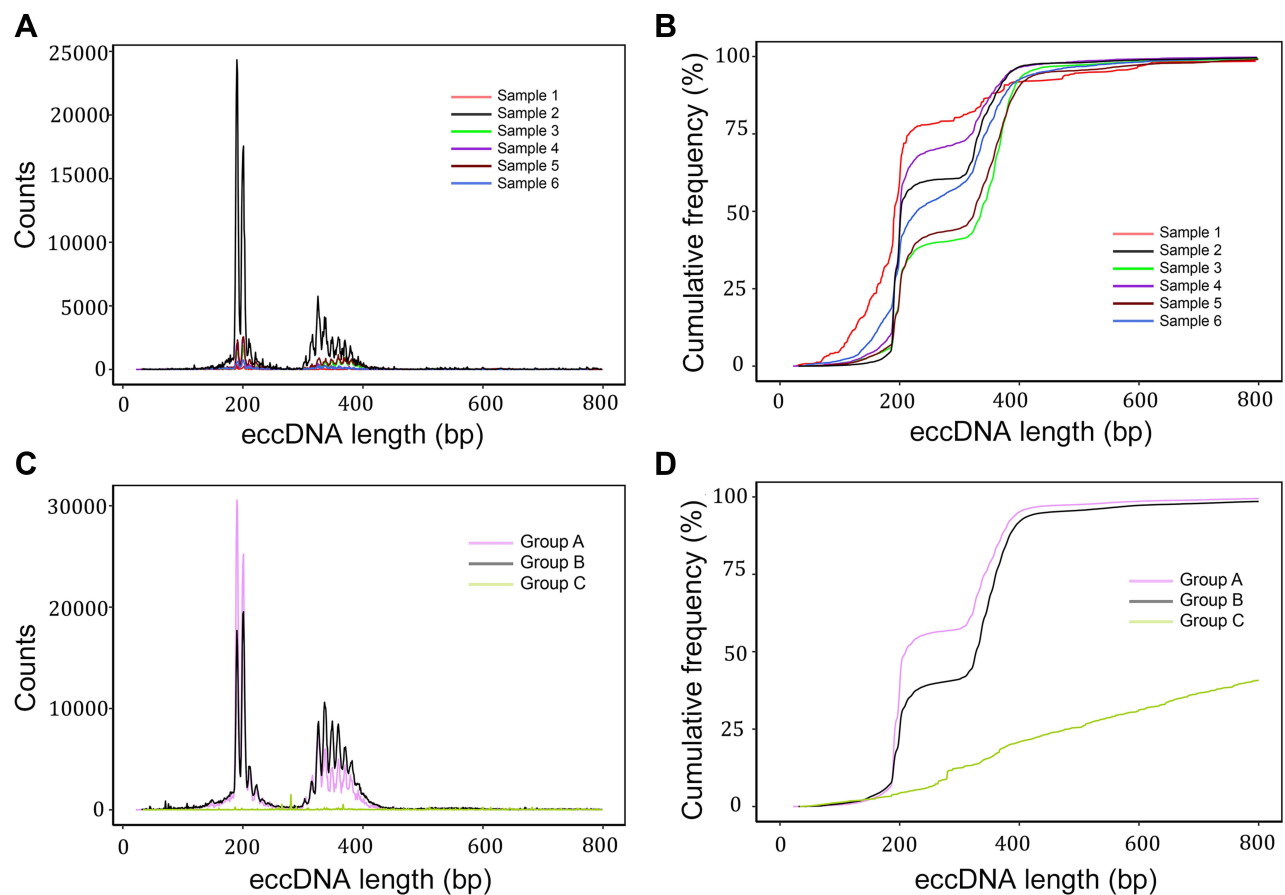
To find differentially expressed eccDNA molecules, we compared the eccDNA expression level between lung adenocarcinoma patients (group A) and healthy people (group B and C). R package DESeq2 (v1.32.0) was used to determine the differentially expressed eccDNA molecules and  $P < 0.05$  were used as the cut-off value. To further screen the possible function of these eccDNA molecules, we used R package clusterProfiler (v4.0.5) to perform the Gene Ontology (GO) function analysis and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the eccDNA molecules hosting genes. For the 9 target eccDNA molecules, we performed survival analysis of the hosting genes of eccDNA molecules in TCGA-LUAD database using R software (v4.1.1). Associations between the overall survival (OS) of TCGA-LUAD patients and gene expression levels were visualized by Kaplan-Meier Plots and compared using Log rank test.

## Results

### General Characteristics of Plasma eccDNA in Lung Adenocarcinoma Patients

We identified a total number of eccDNA from 6, 512 to 455, 371 with split cutoff as 1 in each LUAD patient (details in [Supplementary Table 1](#)). In 6 samples, the detected eccDNA showed bimodality size distribution characteristics in ~191bp and ~320bp ([Figure 2A](#)). The majority (more than 90%) of eccDNA were smaller than 800 bp ([Figure 2B](#)). Based on this observation, we hypothesized that the size distribution of human eccDNA is dominated at two primary peaks in ~191bp and ~320bp ([Figure 2A](#)).

The size distribution of eccDNA in patients with lung adenocarcinoma was compared with that of eccDNA in pregnant women and healthy adults. The size distribution of eccDNA in lung adenocarcinoma is similar to that of pregnant women. Our results also showed that the abundance of eccDNA with size around two primary peaks in lung adenocarcinoma patients was five times more pronounced than that of pregnant women. For the data of healthy person



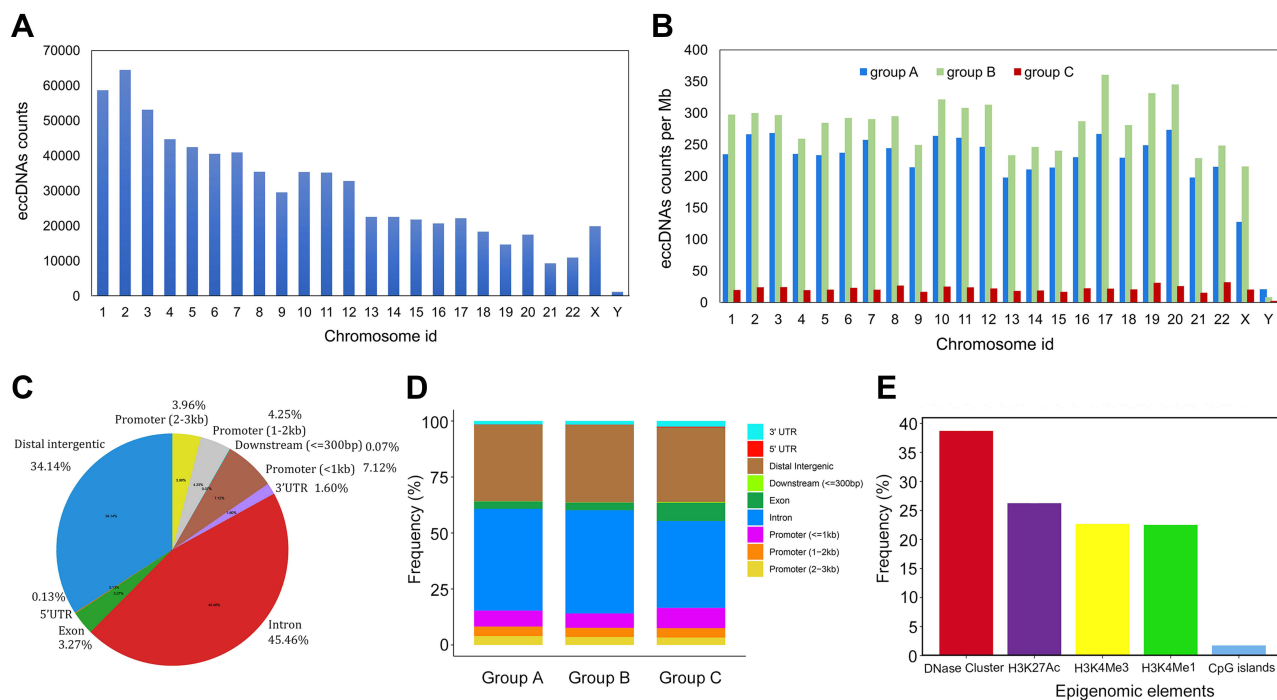
**Figure 2** Size distribution of eccDNA. **(A)** Size distribution of plasma eccDNA in 6 lung adenocarcinoma cases. Plasma samples of 6 lung adenocarcinoma cases were analyzed and displayed with different colors. Plasma eccDNA length in most samples showed similar bimodal peak distribution, indicating that eccDNA in plasma of lung adenocarcinoma patients was quite regular. **(B)** Cumulative frequency plots of plasma eccDNA length from 6 lung adenocarcinoma cases. More than 90% plasma eccDNA in 6 lung adenocarcinoma cases were smaller than 800 bp. **(C)** Size distribution of total eccDNA in plasma of adenocarcinoma patients (group **(A)**), healthy pregnant women (group **(B)**), and healthy adults (group **(C)**). The size of eccDNA is represented by the abscissa, while the number of eccDNA is shown by the ordinate. The size of eccDNA in groups A and B showed distinct bimodal peak distribution at ~190bp and ~330bp, with evident 10-bp periodicity around the two peaks. Plasma eccDNA counts in group C were smaller compared with other groups and no prominent size distribution characteristics was noticed. **(D)** Cumulative frequency plots of plasma eccDNA in adenocarcinoma patients (group **(A)**), healthy pregnant women (group **(B)**), and healthy adults (group **(C)**). Group **(A)** and group **(B)** showed similar bimodal peak distribution.

group, the bimodality size distribution characteristics are not comparable due to its different processes before eccDNA sequencing (Figure 2C and D).

## Genomic Annotation of Plasma eccDNA

To determine if there was a preference in the chromosome origin of eccDNA, we mapped the eccDNA to human genome. Our results indicated that eccDNA fragments can be ubiquitously generated in the genome without evident preference (Figure 3A). Similar results were found in healthy people and healthy gravidas (Figure 3B).

We annotated all eccDNA molecules with different classes of genomic elements (5'UTR, 3'UTR, distal intergenic, exon, intron, downstream and promoters; Figure 3C) The plasma eccDNA in lung adenocarcinoma patients was enriched in introns and distal intergenic regions. Such distribution patterns were in line with healthy adults (group C) and pregnancy women (group B; Figure 3D). All of the detected plasma eccDNA from 6 lung adenocarcinoma patients were mapped to epigenomic elements (CpG Islands, DNase clusters, H3K4Me1, H3K4Me3, and H3K27Ac) using UCSC Table Browser (Figure 3E). The proportion of eccDNA intersected with DNase clusters, H3K27Ac, H3K4Me3, H3K4Me1, CpG islands were 38.77%, 26.31%, 22.74%, 22.58% and 1.81%, respectively.



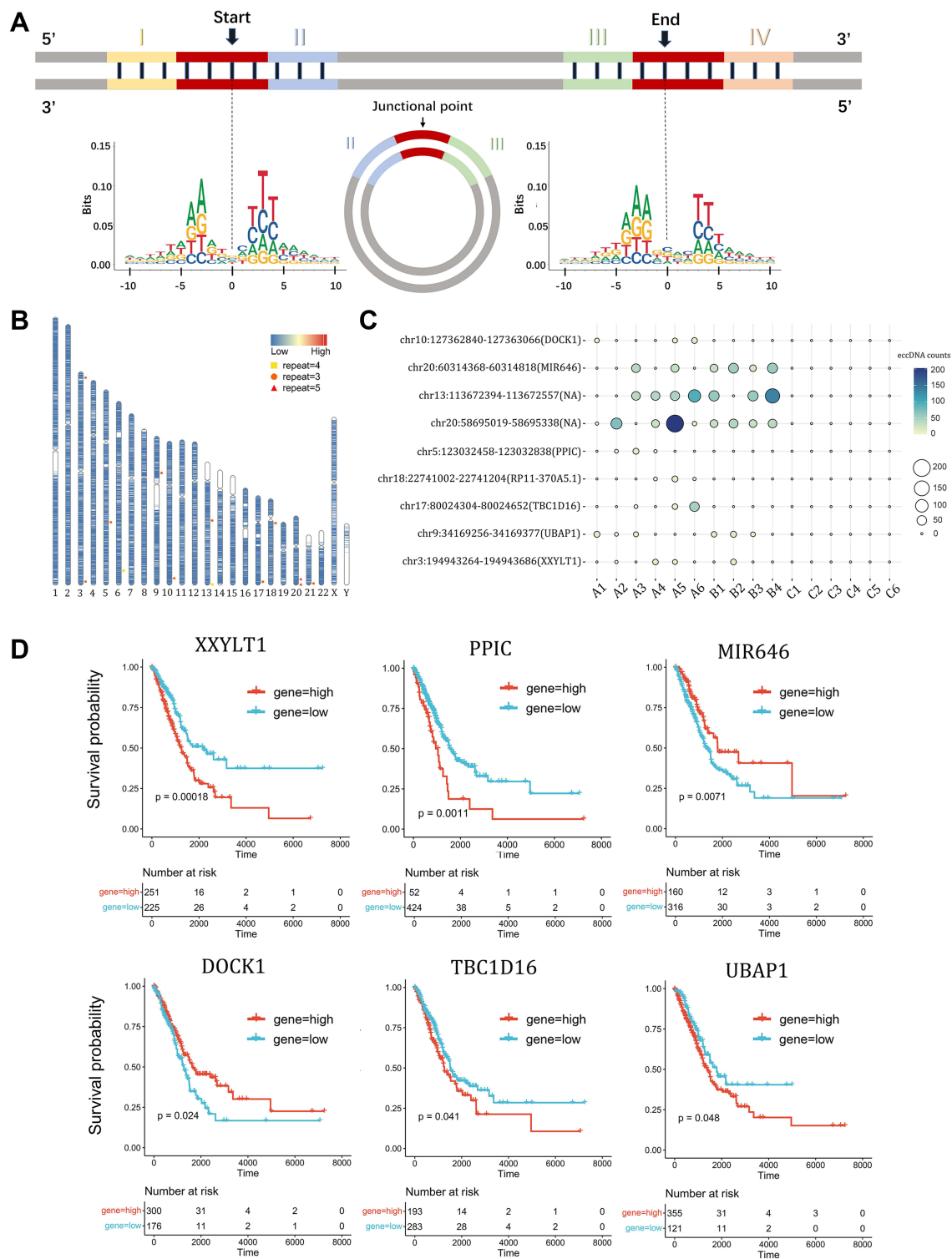
**Figure 3** Chromosome distribution and genomic annotation of eccDNA. **(A)** Genome view of total eccDNA distribution in 6 lung adenocarcinoma patients. The 22 linear autosome pairs and a pair of sex-determining chromosomes are shown on X-axis. Plasma eccDNA molecules originating from each chromosome are shown on Y-axis. **(B)** Genome view of normalized eccDNA distribution in groups A, B, and C. Plasma eccDNA counts per Mb are shown on the Y-axis. Since there were many undetermined genes on the Y chromosome, fewer eccDNA originated from the Y chromosome compared to X chromosome. **(C)** Plasma eccDNA molecules of lung adenocarcinoma patients can be mapped to various genomic regions. 6 lung adenocarcinoma samples were annotated as a group. Higher frequency of eccDNA originating from intron, distal intergenic and promoter regions were observed. 5'-UTR, 5'-untranslated regions; 3'-UTR, 3'-untranslated regions; distal intergenic, regions that do not encode proteins and are far away from gene; downstream ( $\leq 300$ bp), the regions downstream of the gene in a range of 300bp). **(D)** Genomic annotation of eccDNA in groups A, B, and C. Three groups showed similar characteristics with a preference in intron, distal intergenic and promoter regions, indicating that eccDNA generated hotspots in these regions. **(E)** Functional annotation of eccDNA in 6 lung adenocarcinoma samples. The x-axis shows 5 kinds of epigenomic elements, while the Y-axis represents the proportion of this form of modification found in all plasma eccDNA molecules.

## Trinucleotide Motifs Flanking eccDNA Junctions

To determine if there was specific pattern in eccDNA formation, we studied the composition features of the DNA sequence on the left or right sides of eccDNA molecules to determine whether there were certain base sequence characteristics that made chromosomes easier to break, thus supporting the generation of eccDNAs (Figure 4A). The analysis indicated that in the plasma of lung adenocarcinoma patients, the preference for base sequence around the eccDNA breakpoint was homologous to that in pregnant women.

## Differential Expression of eccDNA in LUAD

We observed that eccDNA can be dropped from practically any point on the chromosome after visualizing the loci of the eccDNA chromosome, which is consistent with previous studies.<sup>20,21</sup> Although most eccDNA existed independently in various samples according to the sequencing data, some eccDNA were concurrently present in plasma samples of lung adenocarcinoma. We searched the eccDNA molecules that detected in at least 3 lung adenocarcinoma patients but not identified in samples of healthy people, and 9 kinds of eccDNA molecules were eligible (split reads  $\geq 1$ ). Standardized eccDNA counts of the 9 eccDNA molecules were presented in all three groups (Figure 4B). The detailed information of the 9 eccDNA molecules is shown in the table attached (Supplementary Tables 5 and 6). Based on this premise, plasma eccDNA of healthy normal individuals and healthy pregnant women was compared (Figure 4C). As shown in Figure 4C, the highest frequency of eccDNA reproducibility is 5/6, which was a fragment from chr20: 58695019–58695338. Of the 9 eccDNA molecules, 6 locate on known gene coding region including XXYLT1, PPIC, MIR646, DOCK1, TBC1D16 and UBAP1. We evaluated the prognostic values of these 6 genes and found patients with higher XXYLT1, PPIC,



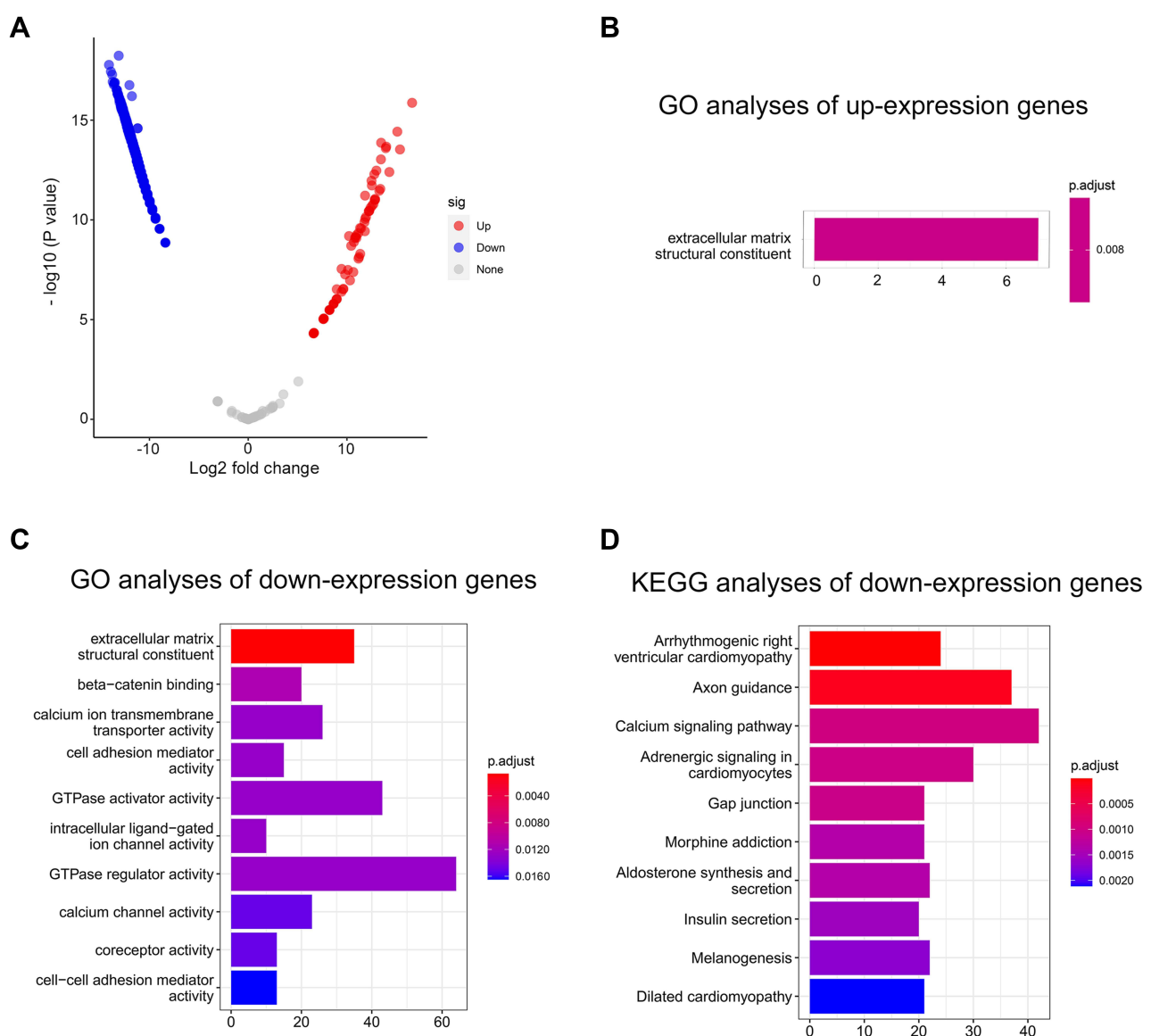
**Figure 4** Identification of eccDNA junctional nucleotide motif patterns and potential biomarkers for the diagnosis of LUAD. **(A)** The four trinucleotide fragments flanking the junctions of eccDNA were named as I, II, III, and IV. eccDNA was clipped at indicated locations (red regions with arrows). 10-bp nucleotide motif sequences flanking the start and end of plasma eccDNA molecules were obtained from the reference genome. The X-axis represents the position while the colored letters represent four kinds of bases of A, T, C, G and Y-axis showed their frequency. Trinucleotide motif patterns of I, II, III, and IV flanking eccDNA junctions with the top 18 frequencies were listed. **(B)** 22 linear autosome pairs and the sexual chromosomes are represented by the ideograms. 24 ideograms are labeled by colored lines, exhibiting the position where eccDNA molecules originated from the chromosome. The eccDNA found in more than three samples are shown by different shapes beside the ideograms. **(C)** Differences in plasma eccDNA molecule expressions between lung adenocarcinoma patients and healthy people. The abscissa shows the lung adenocarcinoma patients (A1 to A6), healthy pregnant female groups (B1 to B4), and healthy adults (C1 to C6). The ordinate represents the eccDNA location on the reference genome and its host-genes names. NA represents eccDNA mapped to the regions without any gene. The detected number of eccDNA is represented by the dotplot size and shade. **(D)** The survival analysis of genes associated with the target eccDNA using the data from the TCGA dataset. Kaplan-Meier plots were drawn and the corresponding risk tables were displayed ( $P < 0.05$ ).

TBC1D16 and UBAP1 were associated with poorer overall survival, while patients with higher MIR646 and DOCK1 were associated with better overall survival (Figure 4D).

## Predictive Biological Function of Differentially Expressed eccDNA in LUAD

Comparing the expression pattern between the plasma of lung adenocarcinoma patients and 10 healthy people (group B and group C), 109 eccDNA hosting genes were up-regulated and 2732 eccDNA hosting genes were down-regulated in lung adenocarcinoma patients. The volcano plot displayed the difference in the expression of eccDNA between lung adenocarcinoma patients and 10 healthy people (group B and group C) (Figure 5A;  $\log_2FC > 1$ ,  $FDR-P < 0.05$ ).

GO (Gene ontology) analysis result of up-regulated eccDNA is shown in Figure 5B and the top ten results of GO analysis and KEGG analysis of down-regulated eccDNA are shown in Figure 5C and D. The GO analysis showed the dominant biological processes of down-regulated eccDNA in LUAD were related to extracellular matrix structural constituent, GTPase activity, transmembrane transporter, cell adhesion, ion channel, gated channel activity, and the cytoskeleton. In addition, the GO analysis



**Figure 5** Differentially expressed eccDNA and functional analysis of their associated genes. **(A)** Volcano plot of significantly differential expressed eccDNA between lung adenocarcinoma patients and healthy people. **(B)** The GO analysis results of host-genes associated with up-regulated eccDNA molecules. **(C)** The top 10 GO analysis results of host-genes associated with down-regulated eccDNA molecules. **(D)** The top 10 KEGG analysis results of host-genes associated with down-regulated eccDNA molecules.



revealed that up-regulated eccDNA in LUAD were related to extracellular matrix structural constituent. KEGG pathway analysis revealed that host-genes on down-regulated eccDNA in LUAD were mainly involved in cardiomyopathy, axon guidance, calcium signaling pathway, hormone secretion, signaling pathways regulating pluripotency of stem cells, Rap1 signaling pathway, cAMP signaling pathway, secretion of digestive juice.

## Discussion

The focus of this research was to identify the similarities and differences between eccDNA in the plasma of healthy people and lung adenocarcinoma patients to determine whether eccDNA has the potential to be used as a cancer biomarker in the early diagnosis of adenocarcinoma. This was the first report with comprehensive evaluation pertaining to factors such as size distribution, GC concentration, genome dispersion, and significant results were yielded demonstrating a difference in terms of type and abundance. These discrepancies could provide a theoretical foundation for the use of eccDNA as biomarker in future clinical studies.

Previous research has demonstrated that the size distribution of eccDNA in pregnant women and fetuses had two primary peaks at ~202 and ~338 bp, respectively. Additionally, both maternal and fetal eccDNA profiles had a 10 bp periodicity, which is consistent with our findings.<sup>19,22</sup> The shift in eccDNA size distribution in cancer patients may be caused by changes in the microenvironment, and the underlying mechanisms warrant further research. According to prior research, eccDNA may play important roles in tumor incidence and progression, tumor heterogeneity and tumor oncogene amplification. The overlapping of eccDNA intersection with epigenomic elements was high, which may be related to the potential of eccDNA.

Paulsen's simulation of microDNA using artificial molecules with known microDNA sequence and structure suggested that microDNA can express tiny regulatory RNA.<sup>23</sup> Coincidentally, of the many eccDNA we screened, the four identified targets with the most biomarker potential are all small in length within 100bp–400bp. This could be a goldilocks range since it strikes a balance between the heightened stability of being short, yet still long enough for regulatory RNA to be transcribed. The expression of regulatory RNA also suggests that eccDNA could play a role in the onset and progression of malignancies, and might be exploited as cancer biomarker.<sup>23</sup>

The functions of eccDNA are not fully understood. We performed GO and KEGG analysis of the host-genes of differentially expressed eccDNA in LUAD. eccDNA hosting genes were significantly enriched in extracellular matrix structural constituent, and some up-regulated host-genes are associated with the progression of LUAD. Meanwhile, the GO and KEGG analysis of down-regulated eccDNA confirmed that extracellular matrix structural, ion transmembrane transport, cell adhesion, endocrine, constituent, and GTPase regulator were highly enriched. Previous studies have showed that many host-genes of differentially expressed eccDNA molecules have been confirmed to be associated with progression of cancer cell. These include PPIC, which presented significant correlations with mitochondrial metabolism, inflammation, folding of multiple proteins, endotoxin signaling and immune response,<sup>24–26</sup> suggesting that host-genes of eccDNA are associated with the occurrence and development of cancer and might become markers for early detection and prognosis of lung cancer.

We were able to successfully screen the unique eccDNA of lung cancer patients, and our results revealed that these unique eccDNA have characteristics such as high GC content, a large number of repetitions, a short sequence and a deficient plasma level, which partially matches the results described by Shibata.<sup>20</sup> These traits spell challenges for the creation of proper experimental conditions since it will be hard to conduct PCR for a given eccDNA. Due to the difficulty of designing primers, it is still challenge before we could verify that the eccDNA we selected can become mature biomarkers for early diagnosis of LUAD.

It is exciting that DNA amplification technologies like rolling circle amplification and assay for transposase-accessible chromatin using sequencing make it possible to analyze eccDNA despite low eccDNA content, although sequencing remains a big challenge.<sup>27,28</sup> Addressing this challenge will be crucial for continued study regarding the clinical viability of eccDNA. Limited to various conditions, we cannot include more samples, as medical technology continues to advance, we predict that the upcoming third-generation sequencing technology will bring opportunities to explore eccDNA as a lung cancer biomarker. In the future, we will continue to apply sophisticated technology to study unique eccDNA sequences of lung adenocarcinoma patients, especially the mechanism of production, distinctive function, and action mechanism.

## Conclusion

In conclusion, we initially reported that LUAD patients have unique eccDNA expression pattern compared to healthy person, and eccDNA has great potential to be used as a biomarker for early detection of lung cancer. In addition, there are various understudied eccDNA in our bodies. Studying the functions of these eccDNA molecules with the help of cutting-edge tools will provide new mechanisms of cancer biology.

## Acknowledgments

This work was supported by the Shanghai Science and Technology Commission (No. 18441905400), Natural Science Foundation of Shanghai (20ZR1434100) and Research on the compliant application for real world data collection, governance and administration of innovation technology in research hospital (No. HNLC2022RWS014). Xiaoqiong Wu and Pu Li are co-first authors for this study.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Disclosure

The authors report no conflicts of interest in this work.

## References

- Cohen S, Yacobi K, Segal D. Extrachromosomal circular DNA of tandemly repeated genomic sequences in *Drosophila*. *Genome Res*. 2003;13(6A):1133–1145. doi:10.1101/gr.907603
- 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. doi:10.1038/s41467-018-03369-8
- Hotta Y, Bassel A. Molecular size and circularity of DNA in cells of mammals and higher plants. *Proc Natl Acad Sci U S A*. 1965;53(2):356–362. doi:10.1073/pnas.53.2.356
- Moller HD, Parsons L, Jorgensen TS, et al. Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci U S A*. 2015;112(24):E3114–E3122. doi:10.1073/pnas.1508825112
- Stanfield SW, Lengyel JA. Small circular DNA of *Drosophila melanogaster*: chromosomal homology and kinetic complexity. *Proc Natl Acad Sci U S A*. 1979;76(12):6142–6146. doi:10.1073/pnas.76.12.6142
- Sunnerhagen P, Sjoberg RM, Karlsson AL, et al. Molecular cloning and characterization of small polydisperse circular DNA from mouse 3T6 cells. *Nucleic Acids Res*. 1986;14(20):7823–7838. doi:10.1093/nar/14.20.7823
- Stanfield SW, Helinski DR. Cloning and characterization of small circular DNA from Chinese hamster ovary cells. *Mol Cell Biol*. 1984;4(1):173–180. doi:10.1128/mcb.4.1.173-180.1984
- Wang T, Zhang H, Zhou Y, et al. Extrachromosomal circular DNA: a new potential role in cancer progression. *J Transl Med*. 2021;19(1):257. doi:10.1186/s12967-021-02927-x
- Wu S, Turner KM, Nguyen N, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature*. 2019;575(7784):699–703. doi:10.1038/s41586-019-1763-5
- Koche RP, Rodriguez-Fos E, Helmsauer K, et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat Genet*. 2020;52(1):29–34. doi:10.1038/s41588-019-0547-z
- Urbanova M, Plzak J, Strnad H, et al. Circulating nucleic acids as a new diagnostic tool. *Cell Mol Biol Lett*. 2010;15(2):242–259. doi:10.2478/s11658-010-0004-6
- Paulsen T, Kumar P, Koseoglu MM, et al. Discoveries of extrachromosomal circles of DNA in normal and tumor cells. *Trends Genet*. 2018;34(4):270–278. doi:10.1016/j.tig.2017.12.010
- 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(11):1749–1759. doi:10.1016/j.celrep.2015.05.020
- Hull RM, King M, Pizza G, et al. Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLoS Biol*. 2019;17(12):e3000471. doi:10.1371/journal.pbio.3000471
- Gresham D, Usaite R, Germann SM, et al. Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the *GAP1* locus. *Proc Natl Acad Sci U S A*. 2010;107(43):18551–18556. doi:10.1073/pnas.1014023107
- Tsai SQ, Nguyen NT, Malagon-Lopez J, et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods*. 2017;14(6):607–614. doi:10.1038/nmeth.4278
- Prada-Luengo I, Krogh A, Maretty L, et al. Sensitive detection of circular DNAs at single-nucleotide resolution using guided realignment of partially aligned reads. *BMC Bioinform*. 2019;20(1):663. doi:10.1186/s12859-019-3160-3

18. Zhu J, Zhang F, Du M, et al. Molecular characterization of cell-free eccDNAs in human plasma. *Sci Rep.* 2017;7(1):10968. doi:10.1038/s41598-017-11368-w
19. Sin S, 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(3):1658–1665. doi:10.1073/pnas.1914949117
20. Shibata Y, Kumar P, Layer R, et al. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science.* 2012;336(6077):82–86. doi:10.1126/science.1213307
21. Sun Z, Ji N, Zhao R, et al. 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
22. Yang H, He J, Huang S, et al. Identification and characterization of extrachromosomal circular DNA in human placentas with fetal growth restriction. *Front Immunol.* 2021;12:780779. doi:10.3389/fimmu.2021.780779
23. 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(9):4586–4596. doi:10.1093/nar/gkz155
24. Paiva RS, Ramos CV, Azenha SR, et al. Peptidylprolyl isomerase C (Ppic) regulates invariant Natural Killer T cell (iNKT) differentiation in mice. *Eur J Immunol.* 2021;51(8):1968–1979. doi:10.1002/eji.202048924
25. Gao YF, Zhu T, Mao CX, et al. PPIC, EMP3 and CHI3L1 are novel prognostic markers for high grade glioma. *Int J Mol Sci.* 2016;17(11):1808. doi:10.3390/ijms17111808
26. Chapman DC, Stocki P, Williams DB. Cyclophilin C participates in the US2-mediated degradation of major histocompatibility complex class I molecules. *PLoS One.* 2015;10(12):e145458. doi:10.1371/journal.pone.0145458
27. Cao X, Wang S, Ge L, et al. Extrachromosomal circular DNA: category, biogenesis, recognition, and functions. *Front Vet Sci.* 2021;8:693641. doi:10.3389/fvets.2021.693641
28. 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(20):a2489. doi:10.1126/sciadv.aba2489

International Journal of General Medicine

Dovepress

## Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>