OXFORD

*Briefings in Bioinformatics*, 2024, **25(2)**, 1–8
https://doi.org/10.1093/bib/bbae034
**Problem Solving Protocol**

# eccDNA-pipe: an integrated pipeline for identification, analysis and visualization of extrachromosomal circular DNA from high-throughput sequencing data

Minghao Fang†, Jingwen Fang†, Songwen Luo, Ke Liu, Qiaoni Yu, Jiaxuan Yang, Youyang Zhou, Zongkai Li, Ruoming Sun, Chuang Guo and Kun Qu

Corresponding authors. Kun Qu, E-mail: qukun@ustc.edu.cn; Chuang Guo, E-mail: gchuang@ustc.edu.cn; Jingwen Fang, E-mail: fjingwen@ustc.edu.cn
†These authors contributed equally to this work.

## Abstract

Extrachromosomal circular DNA (eccDNA) is currently attracting considerable attention from researchers due to its significant impact on tumor biogenesis. High-throughput sequencing (HTS) methods for eccDNA identification are continually evolving. However, an efficient pipeline for the integrative and comprehensive analysis of eccDNA obtained from HTS data is still lacking. Here, we introduce eccDNA-pipe, an accessible software package that offers a user-friendly pipeline for conducting eccDNA analysis starting from raw sequencing data. This dataset includes data from various sequencing techniques such as whole-genome sequencing (WGS), Circle-seq and Circulome-seq, obtained through short-read sequencing or long-read sequencing. eccDNA-pipe presents a comprehensive solution for both upstream and downstream analysis, encompassing quality control and eccDNA identification in upstream analysis and downstream tasks such as eccDNA length distribution analysis, differential analysis of genes enriched with eccDNA and visualization of eccDNA structures. Notably, eccDNA-pipe automatically generates high-quality publication-ready plots. In summary, eccDNA-pipe provides a comprehensive and user-friendly pipeline for  customized analysis of eccDNA research.

***Keywords***: eccDNA; pipeline; short-read sequencing; long-read sequencing; eccDNA-pipe

## INTRODUCTION

Extrachromosomal circular DNA (eccDNA) refers to circular DNA molecules originating from chromosomes but are not necessarily independent of them, varying in size from less than 100 bp to several megabases [1]. With rapid advancements in sequencing and imaging technologies, eccDNA has been widely detected in eukaryotes ranging from yeast to human and various cancerous and non-cancerous human tissues [2–5]. By driving oncogene overexpression and fostering genetic diversity, eccDNA plays a crucial role in tumor evolution and resistance to therapies [6–9]. Additionally, eccDNA can influence spermatogenesis and act as innate immunostimulants in normal tissue [10, 11]. Due to its structural diversity, functional versatility and variable abundance, investigations into eccDNA shed light on the molecular mechanisms underlying DNA repair pathways, genomic instability and gene expression regulation. Therefore, comprehensive identification and investigation of eccDNA functions have substantial clinical implications including cancer diagnosis and treatment.

The experimental techniques for eccDNA identification can be classified based on the sequencing platform used (short-read or long-read sequencing) and with or without the utilization of rolling cycle amplification (RCA). The utilization of RCA can be categorized into non-circular enrichment methods, such as whole-genome sequencing (WGS) [12], and circular enrichment methods, such as Circle-seq [13] and Circulome-seq [14]. Consequently, various software tools have been developed to identify eccDNA from different experimental techniques. These tools include AmpliconArchitecture (AA) [15] for WGS data using short-read sequencing, Circle-Map [16], Circle_finder [17], ecc_finder [18] and ECCsplorer [19] for processing Circle-seq and Circulome-seq data obtained from short-read sequencing, as well as NanoCircle [20], eccDNA_RCA_nanopore [11], CReSIL

**Minghao Fang** is a graduate student from Qu's Lab.
**Jingwen Fang** is the chief executive officer of HanGen Biotech.
**Songwen Luo** is a graduate student from Qu's Lab.
**Ke Liu** is a graduate student from Qu's Lab.
**Qiaoni Yu** is a graduated Ph.D. from Qu's Lab.
**Jiaxuan Yang** is a staff of HanGen Biotech.
**Youyang Zhou** is a staff of HanGen Biotech.
**Zongkai Li** is a graduate student from Qu's Lab.
**Ruoming Sun** is a graduate student from Qu's Lab.
**Chuang Guo** is an associate professor at Division of Life Sciences and Medicine, USTC.
**Kun Qu** is currently a professor of Genomics and Bioinformatics at Division of Life Sciences and Medicine, USTC. From 2010 to 2016, he was a Bioinformatics Scientist and then the Director of Bioinformatics at Stanford University, USA.

**Received:** October 20, 2023. **Revised:** December 22, 2023. **Accepted:** January 11, 2024

[21] applicable to WGS and Circle-seq generated by long-read sequencing. However, the variability in eccDNA detection bias by these methods leads to inconsistencies. Typically, studies combining several techniques is necessary to achieve a thorough identification and analysis of eccDNA in specimens [22–24].

While existing tools specialize in particular techniques and primarily address eccDNA identification in upstream analysis, researchers encounter challenges in effectively handling eccDNA data and conducting systematic analysis due to the absence of a universal and comprehensive analysis pipeline for both upstream and downstream analysis. The circdna(nf-core) [25] is a simple pipeline for eccDNA identification in upstream analysis;however, it does not incorporate the latest long-read sequencing eccDNA identification algorithms in upstream analysis and none of the downstream analysis software. Besides, eccDNA experiments provide rich information on eccDNA; however, their chromosomal elements, circular characteristics and differentially expressed genes (DEGs) associated with them were barely illustrated by any of the known analytical pipelines. Therefore, a pipeline for more thorough and systematic exploration of the eccDNA source data is necessary.

Here, we introduce eccDNA-pipe, a comprehensive and highly integrated pipeline featuring both upstream and downstream analysis. In the upstream analysis, eccDNA-pipe provides quality control and eccDNA identification. AA has been developed for the identification of tumor ecDNA, which represents longer sequence lengths of eccDNA. The fine structure of focally amplified regions was reconstructed using WGS short-read sequencing data, which have been extensively applied in analyzing datasets from multiple tumors [7, 12]. Circle-Map and CReSIL have emerged as the most effective pipelines for eccDNA detection through both short-read and long-read sequencing data [26]. The eccDNA identification modules incorporated in eccDNA-pipe include AA, Circle-Map and CReSIL, which are by far the best algorithms for either short-read or long-read sequencing eccDNA analysis. In the downstream analysis, eccDNA-pipe supports the automated extraction of upstream identified eccDNA results for eccDNA length distribution analysis, differential gene analysis related to enriched eccDNA and visualization of eccDNA structures. Moreover, the eccDNA pipeline adopts a modular programming approach and allows for input and downstream analysis of eccDNA results from other eccDNA, identification software tools.

# MATERIALS AND METHODS

## Workflow and algorithm

eccDNA-pipe requires Circle-Map, AmpliconArchitecture, CReSIL, trim_galore [27], HOMER [28], fastQC [29], Python and R environment installed on LINUX/MAC OS platform, as well as the corresponding genome references and relative R package limma [30], edgeR [31], DESeq2 [32], Circlize [33] and clusterProfiler [34]. The overall workflow, as depicted in Figure 1, includes all the steps of the eccDNA pipeline, which include two main sections: (1) upstream analysis and (2) downstream analysis. The eccDNA pipeline supports parallel computing options for analyzing large datasets on computer clusters, and users also have the flexibility to execute specific sections of the pipeline. For more detailed information on the step-by-step procedure and usage guidelines, please visit https://github.com/QuKunLab/ecc_pipe.

## Upstream analysis: quality control

In the quality control step, the pipeline begins with the raw sequencing output (FastQ files) and utilizes trim_galore to perform read quality control and adapter trimming. The clean reads are then retained for subsequent sequence analysis. The specific parameters for trim_galore are as follows: -q 30 -phred33 –stringency 3 –length 20 -e 0.1.

## Upstream analysis: eccDNA identification

In the eccDNA identification step, to achieve optimal eccDNA identification performance across different sequencing libraries, the pipeline integrates AmpliconArchitecture, Circle-Map and CReSIL for all sequencing-type data generated by short-read sequencing or long-read sequencing. In the integration of Circle-Map, the pipeline incorporates the Realign functionality. AmpliconArchitecture integrates the python3 version of AmpliconSuite-pipeline, and it allows adjustment of the parameters cngain and cnsize, with default values set at 4 and 10 000. Both Circle-Map and CReSIL utilize default parameters. This integration is facilitated through the snakemake framework [35], enabling eccDNA detection for all types of sequencing data.

## Downstream analysis

In the downstream analysis step, the pipeline supports the standardization of output files from different methods for accessing eccDNA and allows manual input of output files from other tools. Additionally, all downstream analysis results, including PDF analysis reports and corresponding data, are uniformly output to a designated directory. This step involved three functionalities: distribution analysis (Distribution), identification of DEGs based on eccDNA count and eccDNA structure visualization (Visualization).

For each sample, the Distribution functionality generates the fragment size distribution and chromosome distribution. It further annotates and quantifies chromosomal elements with annotatePeaks.pl in the HOMER suite. Additionally, leveraging publicly available annotation files from the eccDNA Atlas [36], it annotates and quantifies enhancers, super-enhancers, single nucleotide polymorphisms (SNPs) and expression quantitative trait locis (eQTLs). Moreover, it automatically generates HTML reports using Jinja2 [37].

The DEG of eccDNA functionality requires a .txt file containing grouping information for each sample. The script exports an N X M data matrix D, where N represents the total number of genes on eccDNA and M is the number of samples. Each element Di,j denotes the count of gene i in sample j. The annotation of genes provides two modes: complete gene annotation (genes on the eccDNA fragments that named gene) and gene circularization rate (eccDNA fragments on the genes that named region), with an adjustable annotation ratio parameter for dynamic adjustment. To assess differential genes between different samples, the eccDNA pipeline applies DESeq2, edgeR and limma for normalization and differential analysis on the raw gene matrix, providing a volcano plot. Users have the flexibility to customize the *P*-value and fold-change for creating a tailored volcano plot to visualize the results of DEGs. For the set of DEGs, it performs Gene Ontology (GO) functional enrichment annotation and gene set enrichment analysis (GSEA) using clusterProfiler.

In the Visualization functionality, eccDNA circular visualization is implemented based on Circlize. The structure of eccDNA is predicted by upstream identification algorithms. The visualization of eccDNA structures is facilitated by the Circlize R package. This functionality allows for the clear representation and annotation of eccDNA regions along with their associated genes. All the aforementioned results are generated in a well-organized folder, facilitating easy sharing and storage.

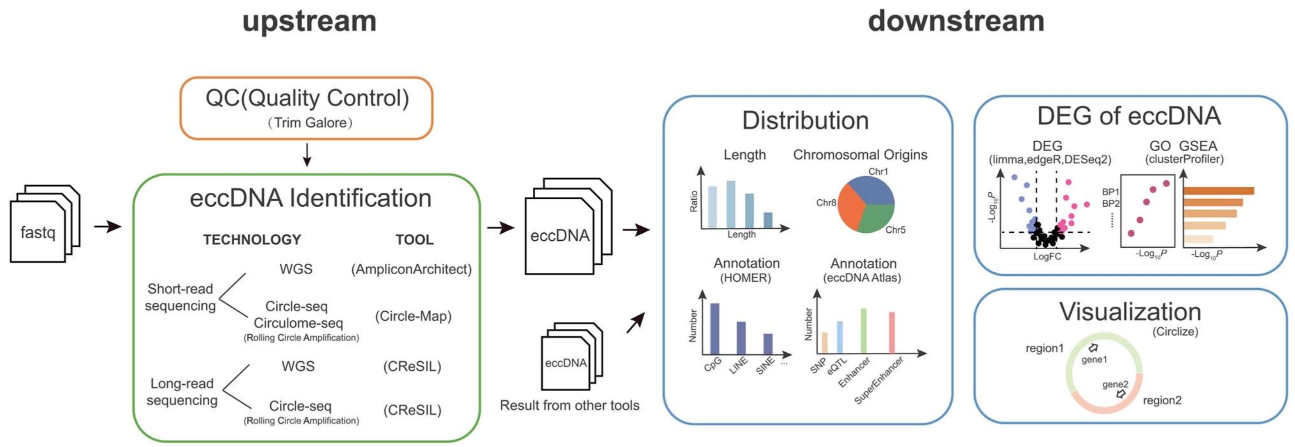

**Figure 1.** Overview of eccDNA-pipe. Users can choose appropriate tools for eccDNA identification from the raw FastQ data during upstream analysis. Additionally, they can conduct fundamental downstream analysis on the eccDNA result files. These analyses include eccDNA length distribution analysis, differential gene analysis related to enriched eccDNA and visualization of eccDNA structures.

## Time and memory comparison

To assess the impacts of different execution strategies on the computing resources consumed by using eccDNA-pipe or the corresponding individual tool to identify eccDNA, we have utilized a computer cluster equipped with two Intel Xeon Scale 6248 CPUs (2.5 GHz, 320 CPU cores), 384 GB of DDR4 memory and 2 TB of AEP memory. We have used the exclusive node with 40 threads simultaneously in parallel to calculate the computing time and memory consumed by each method including Circle-Map, CReSIL and AA.

## Source dataset

Circle-seq short-read sequencing data of human muscle or human blood was downloaded from GEO: PRJNA419440 (SRR6315393-SRR6315434) [38]. Circle-seq long-read sequencing of four human multiple myeloma cancer cell line data (EJM, JJN3, APR1_rep1, APR1_rep2) and three mouse cell line data (5TMG1 multiple myeloma cells and E0771 breast cancer cells, MLOY4 osteocyte-like cells) were downloaded from GEO: PRJNA806866 [21]. WGS short-read sequencing data of GBM39 was downloaded from GEO: PRJNA506071 [8]. Circulome-seq short-read sequencing data of medulloblastoma (MB) were downloaded from GEO: GSE205178 [39].

# RESULTS

## Quality control and eccDNA identification

We collected 42 human muscle or human blood Circle-seq short-read sequencing data [38], 4 human multiple myeloma cancer cell line data (EJM, JJN3, APR1_rep1, APR1_rep2) and 3 mouse cell line data (5TMG1 multiple myeloma cells, E0771 breast cancer cells, MLOY4 osteocyte-like cells) in Circle-seq long-read sequencing [21], GBM39 WGS short-read sequencing data [8] and MB Circulome-seq short-read sequencing data [39] to test the eccDNA-pipe. For the upstream analysis, we modified parameters in the config file and executed trim_galore to remove adapters from the data. In a similar manner, we adjusted and refined the parameters to execute Circle-Map for circle-seq short-read sequencing data [38] and MB Circulome-seq short-read sequencing data [39]. Subsequently, we executed AmpliconArchitecture on GBM39 WGS short-read sequencing data [8] and executed CReSIL for EJM, JJN3, APR1_rep1, APR1_rep2, 5TMG1, E0771 and MLOY4 circle-seq long-read sequencing data [21]. Finally, we obtained the identification of eccDNA results for each tool through eccDNA-pipe upstream analysis. We also conducted time and memory comparisons between eccDNA-pipe and individual algorithms (Supplementary Figure S1A and B, Supplementary Table S1). The results indicate that eccDNA-pipe does not add additional memory and time, and it offers the convenience of setting parallelism with uniform parameters across multiple tools.

## Distribution

The eccDNA-pipe facilitates Distribution analysis of upstream-generated eccDNA result files, and the results are presented in a user-friendly and convenient manner. Using human muscle circle-seq short-read sequencing data (SRR6315430) [38] as an example, the length distribution of eccDNA type and count can be displayed (Figure 2A), wherein the data primarily fall within the range of 100–1000 base pairs. Furthermore, the distribution of eccDNA was proportional across the original chromosomes, with Chr1, Chr2 and Chr3 showing relatively larger proportions (Figure 2B). Additionally, eccDNA-pipe was combined with annotatePeaks.pl in the HOMER suite to uncover the distribution of chromosomal elements. In muscle tissue eccDNA, introns exhibited the highest prevalence, while repetitive elements were prominently enriched in the proximity of SINE and LINE (Figure 2C). Moreover, relevant annotation files were obtained from the eccDNA Atlas to uniformly annotate eccDNA fragments and generate corresponding overlap results for SNPs, Super Enhancers, Enhancers and eQTLs. The analysis indicated a higher prevalence of eccDNA overlap with eQTLs in muscle tissue (Figure 2D). Finally, the eccDNA-pipe automated the generation of HTML reports using Jinja2, which incorporated interactive plotly visualizations and included all annotation results from the Distribution analysis (Supplementary Figure S2).

For multiple samples, the eccDNA-pipe offers a comparative analysis based on the results from each sample. We compared the number of identified eccDNA using eccDNA-pipe with the published gold-standard data from EJM, JJN3, APR1_rep1, APR1_rep2, 5TMG1, E0771 and MLOY4 circle-seq long-read sequencing data [21]. The results suggest that the eccDNA-pipe identification outcomes closely correspond to the published data, affirming the reliability of eccDNA-pipe (Figure 3A). We presented the length distribution of eccDNA (Figure 3B, Supplementary Figure S3A), the

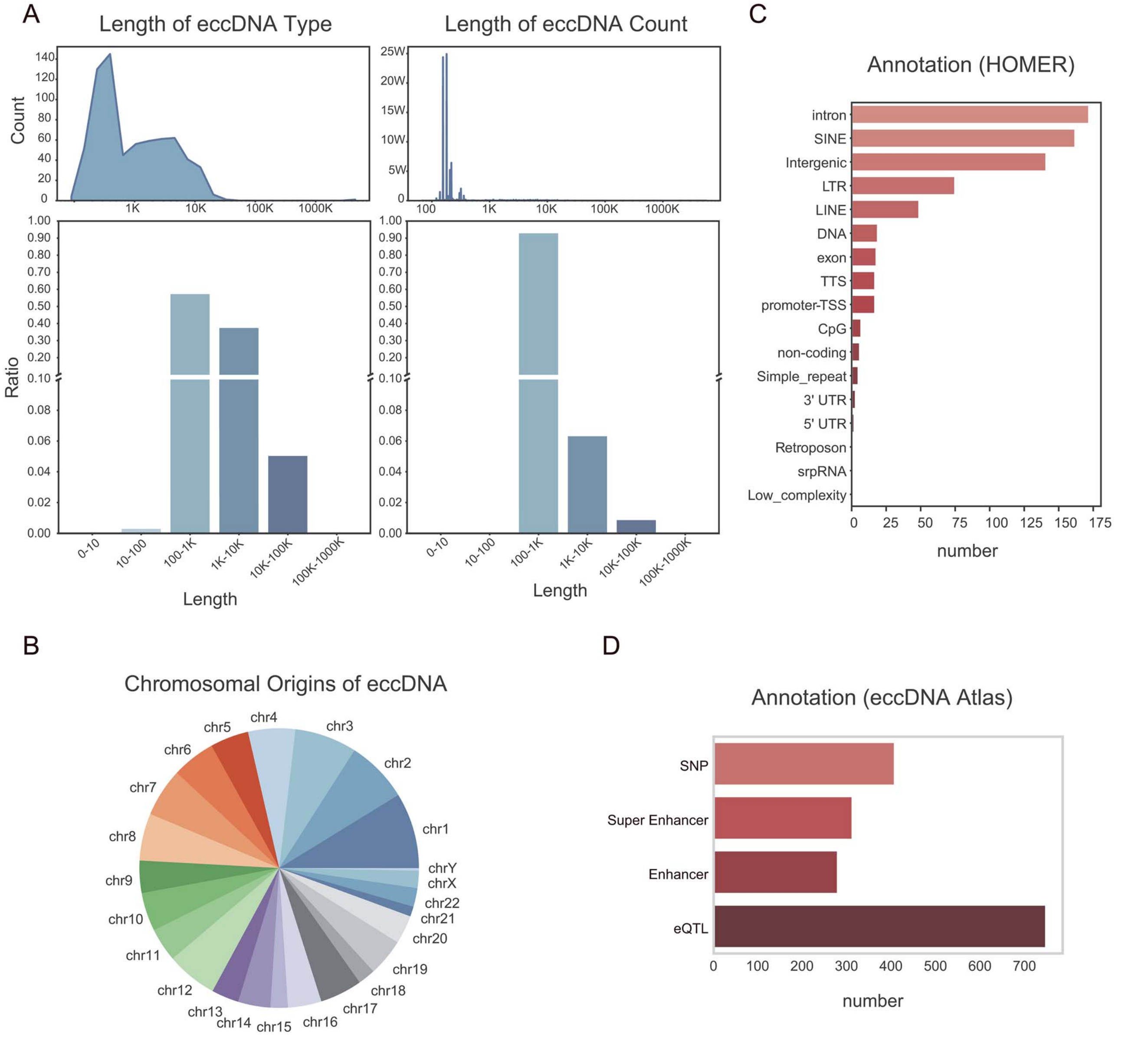

**Figure 2.** Examples of eccDNA-pipe Distribution outputs (SRR6315430 human muscle circle-seq short-read sequencing data). (**A**) Barplot and histogram visualization for eccDNA type/count in length (type: each type of eccDNA counts as 1; count: each type of eccDNA is calculated based on its count number). (**B**) Pie chart illustrating eccDNA chromosomal origins. (**C**) Barplot showing the number of chromosomal elements associated with eccDNA using HOMER annotation. (**D**) Barplot showing the number of regulatory elements (e.g. SNPs, SuperEnhancers, Enhancers, eQTLs) associated with eccDNA using eccDNA Atlas annotation.

chromosomal distribution of eccDNA (Figure 3C) and the chromosomal element distribution of eccDNA (Supplementary Figure S3B) from EJM, JJN3, APR1_rep1, APR1_rep2, 5TMG1, E0771 and MLOY4 circle-seq long-read sequencing data. In this comparison, JJN3 shows a higher proportion of fragments below 500 base pairs compared to other samples. Moreover, we applied eccDNA-pipe to 42 human muscle or human blood Circle-seq short-read sequencing data [38] to show its analytical capabilities for processing larger-scale datasets. We presented the length distribution of eccDNA (Figure 3D, Supplementary Figure S3C) and the chromosomal distribution of eccDNA (Supplementary Figure S3D) from 42 human muscle or human blood Circle-seq short-read sequencing data. It consistently exhibits a higher proportion of fragments below 1000, aligning with the captured length distribution of Circle-seq.

## Visualization

The eccDNA-pipe offers the ability to perform circular visualization of user-defined selected eccDNA. For instance, using GBM39 WGS short-read sequencing data [8], based on the upstream analysis of eccDNA-pipe, we were able to reproduce the results containing *EGFR* eccDNA. The AA of eccDNA-pipe directly displayed the prediction results for this Amplicon (Supplementary Figure S3E). By utilizing the Visualization, we observed a clear enrichment of the *EGFR* gene in the middle of the two segments of this eccDNA (Figure 3E).

## DEG of eccDNA

The eccDNA-pipe provides the functionality to annotate eccDNA fragments at the gene level and perform differential gene analysis based on grouping. Compared with other individual algorithms

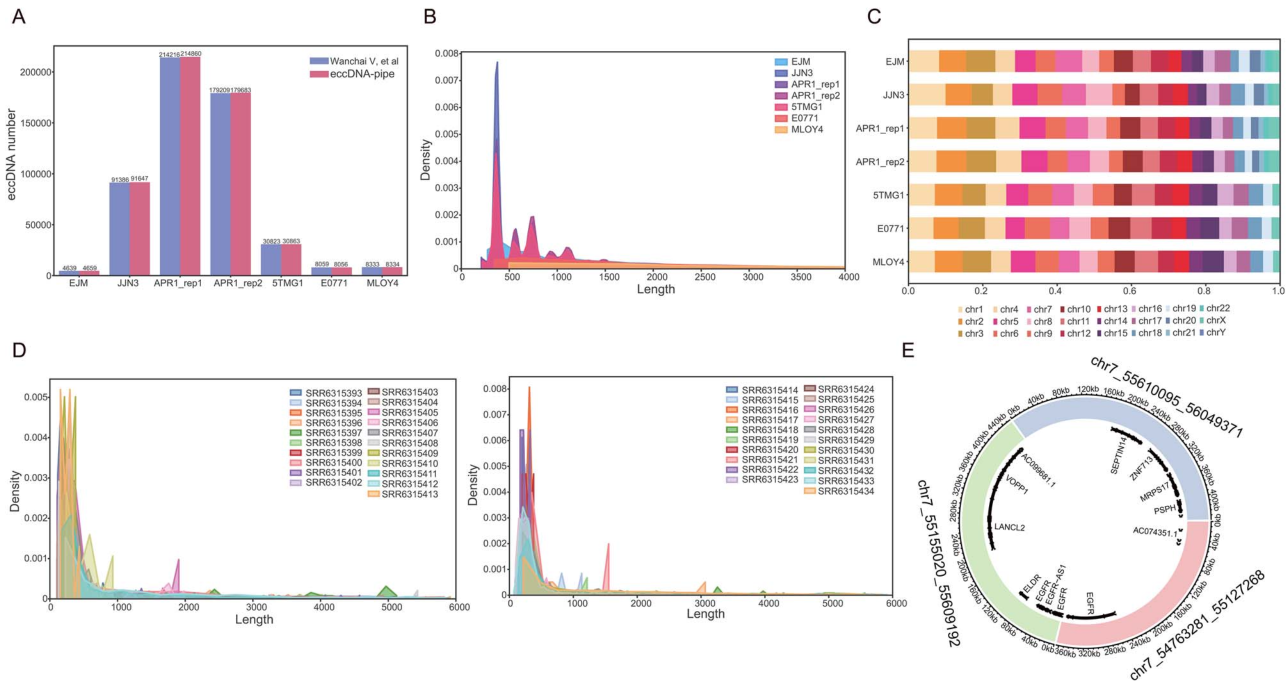

**Figure 3.** The application of eccDNA-pipe in various datasets. (**A**) The comparison of number of eccDNAs discovered by eccDNA-pipe and the original study using CReSIL from the mouse/human cell line long-read sequencing data. (**B**) The length distribution of the eccDNAs (≤4000 bp) discovered by eccDNA-pipe from mouse/human cell line long-read sequencing data. (**C**) The chromosomal distribution of the eccDNAs discovered by eccDNA-pipe from mouse/human cell line long-read sequencing data. (**D**) The length distribution of eccDNAs (≤6000 bp) discovered by eccDNA-pipe from 42 short-read sequencing datasets (SRR6315393-SRR6315434). (**E**) The circular visualization of *EGFR* eccDNA by eccDNA-pipe visualization in GBM39 WGS short-read sequencing data.

and existing pipelines, eccDNA-pipe is the first to develop compatible DEG functionality (Supplementary Figure S4). Taking MB Circulome-seq short-read sequencing data [39] as an example, we identified a higher quantity of eccDNA in the tumor group compared to the normal group using Circle-Map in eccDNA-pipe (Figure 4A). It aligns with the conclusions in the original literature. We also presented the chromosomal distribution of eccDNAs of tumor and normal samples (Figure 4B). As Circulome-seq short-read sequencing data consist of short reads, we selected an overlap ratio of 0.5 in region mode to calculate the number of eccDNAs on genes. Subsequently, we performed differential gene analysis, GO annotation and GSEA based on the raw gene X sample matrix using limma and clusterProfiler. The results revealed that in the tumor group, there was enrichment of eccDNA in the top two enriched pathways of glutamate receptor signaling pathway (*GNAQ, GRIK3, GRM1, GRIA3, TRPM1, GRIK5, KCNB1*) and cell–cell adhesion via plasma–membrane adhesion molecules (*PCDHA1, PCDHA2, PCDHA3, TENM3, CDH20*) (Figure 4C–E, Supplementary Tables S2 and S3).

## DISCUSSION

Different single algorithms present various challenges in terms of parallel parameters, output files and environmental compatibility when handling datasets generated from distinct sequencing platforms and experimental approaches. This diversity can result in an inconvenient user experience for researchers. Furthermore, single algorithm does not support systematic downstream analysis, making it challenging for users to delve into deeper interpretation of eccDNA results. In contrast, eccDNA-pipe uniformly integrates CReSIL, AA and Circle-Map using the Snakemake framework for upstream analysis, providing users with a convenient solution for processing large-scale data with these algorithms. For downstream analysis, a unified processing and analysis workflow was established for eccDNA-pipe through Python programming. This allows users to systematically conduct both upstream and downstream analyses using a single pipeline. In summary, eccDNA-pipe offers a pioneering solution by providing comprehensive and integrated analysis capabilities for both upstream and downstream processes. However, eccDNA-pipe still has some limitations and areas for improvement.

The eccDNA-pipe currently offers downstream analyses including eccDNA length distribution, differential gene analysis related to enriched eccDNA and visualization of eccDNA structures. Additional downstream analyses of eccDNAs, such as structural variation (SV), single-nucleotide variation (SNV) and enhancer hijacking analysis [40] as well as algorithms for single-cell eccDNA analysis, will be developed and integrated into eccDNA-pipe.

Long-read sequencing data play a crucial role in identifying complex eccDNAs due to extended read lengths. Currently, various algorithms, including FLED [41] and Decoil [42], have been developed to identify more reliable eccDNAs from long-read sequencing data. We plan to incorporate these newly developed algorithms into our pipeline.

Different eccDNA identification algorithms are designed to cater to distinct species. The eccDNA-pipe addresses this challenge by offering an application programming interface (API) and tutorial, enabling users to customize and modify the species if the underlying algorithm supports such modification. For instance, the developers of AA currently provide support for human, mouse and human–viral hybrid. Nevertheless, eccDNA-pipe remains committed to ongoing maintenance and updates.

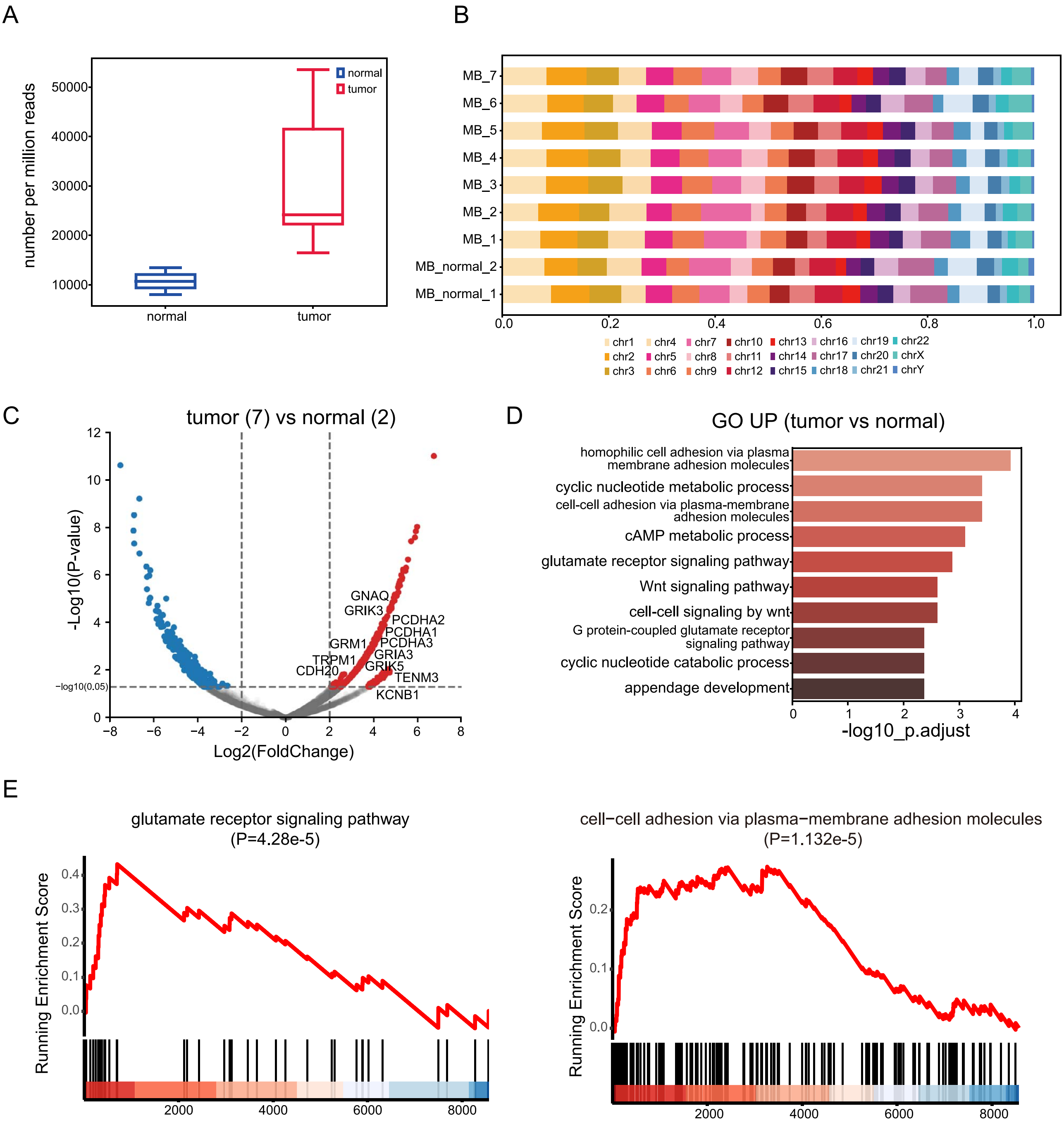

**Figure 4.** The application of eccDNA-pipe to MB Circulome-seq short-read sequencing data. (**A**) Comparison of the number of eccDNAs identified between tumor and normal samples. (**B**) Chromosomal distribution of eccDNAs in tumor and normal samples. (**C**) The volcano plot of eccDNA-pipe DEG according to the eccDNA results (tumor versus normal). (**D**) The GO result of DEG associated with the eccDNA upgenelist (tumor versus normal) by clusterProfiler. (**E**) The GSEA results of the top two enriched pathways of the DEG (tumor versus normal) annotated in eccDNA (i.e. glutamate receptor signaling pathway, cell–cell adhesion via plasma–membrane adhesion molecules).

Although there are inconsistencies in the definitions of ecDNA and eccDNA used by various studies [12, 22, 43], it does not affect the results obtained from eccDNA-pipe. In general, we take those eccDNAs obtained from AA with length >10 kb as ecDNAs.

In summary, eccDNA-pipe is an automated and efficient pipeline designed for the comprehensive analysis of eccDNA. This user-friendly tool transforms the analysis results into publication-quality plots and provides a general pipeline for researchers to manipulate eccDNA results obtained from various sequencing methods. By simplifying data exploration of eccDNA information, eccDNA-pipe contributes to the advancement of eccDNA research.

### Key Points

- The eccDNA-pipe is a comprehensive pipeline that integrates well-established algorithms included Circle-Map, AmpliconArchitecture and CReSIL to facilitate eccDNA identification. Additionally, eccDNA-pipe extends its utility by providing downstream analysis capabilities for eccDNA result files. These encompass eccDNA length distribution analysis, differential gene analysis related to enriched eccDNA and visualization of eccDNA structures.

- Using the eccDNA-pipe, researchers can easily install tool environments and utilize multiple threads to accelerate the parallel processing of multiple samples. Moreover, they can annotate chromosomal elements and genes related to enriched eccDNA by performing differential gene analysis (differentially expressed gene) of eccDNA count on multiple grouped samples, enabling exploration of the biological relevance of eccDNA.
- The eccDNA-pipe automatically converts these results into intuitive plots at publication quality based on statistical analysis.

## SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/bib.

## ACKNOWLEDGEMENTS

We thank the USTC supercomputing center and the School of Life Science Bioinformatics Center for providing computing resources for this project.

## FUNDING

This work was supported by the National Key R&D Program of China (2020YFA0112200 and 2022YFA1303200 to K.Q.), the National Natural Science Foundation of China grants (T2125012 and 91940306 to K.Q.; 32270978 to C.G.; 32100457 to J.F.), CAS Project for Young Scientists in Basic Research YSBR-005 (to K.Q.), Anhui Province Science and Technology Key Program (202003a07020021 to K.Q.) and the Fundamental Research Funds for the Central Universities (YD2070002019, WK9110000141 and WK2070000158 to K.Q.).

## AUTHOR CONTRIBUTIONS

K.Q., C.G. and J.F. conceived the project. J.F. and M.F. designed the framework and performed data analysis with the help from Q.Y., J.Y., Y.Z., S.L. and K.L. M.F. and C.G. wrote the manuscript with input from all authors. C.G. and J.F. supervised the entire project. All authors read and approved the final manuscript.

## SOFTWARE AVAILABILITY

eccDNA-pipe is freely available at https://github.com/QuKunLab/ecc_pipe.

## DATA AVAILABILITY

There are no new data associated with this article. Published datasets used in this study: Circle-seq short-read sequencing data of human muscle or human blood was downloaded from GEO: PRJNA419440 (SRR6315393-SRR6315434) [38]. Circle-seq long-read sequencing of four human multiple myeloma cancer cell line data (EJM, JJN3, APR1_rep1, APR1_rep2) and three mouse cell line data (5TMG1 multiple myeloma cells and E0771 breast cancer cells, MLOY4 osteocyte-like cells) were downloaded from GEO: PRJNA806866 [21]. WGS short-read sequencing data of GBM39 was downloaded from GEO: PRJNA506071 [8]. Circulome-seq short-read sequencing data of medulloblastoma (MB) were downloaded from GEO: GSE205178 [39]. The source code of eccDNA-pipe was released at https://github.com/QuKunLab/ecc_pipe.

## CONFLICT OF INTEREST

Jingwen Fang is the chief executive officer of HanGen Biotech. The other authors declare that they have no conflict of interest.

## REFERENCES

1. Ling X, Han Y, Meng J, *et al.* Small extrachromosomal circular DNA (eccDNA): major functions in evolution and cancer. *Mol Cancer* 2021;**20**:113.
2. Noer JB, Horsdal OK, Xiang X, *et al.* Extrachromosomal circular DNA in cancer: history, current knowledge, and methods. *Trends Genet* 2022;**38**:766–81.
3. Wu S, Bafna V, Chang HY, Mischel PS. Extrachromosomal DNA: an emerging hallmark in human cancer. *Annu Rev Pathol* 2022;**17**: 367–86.
4. Libuda DE, Winston F. Amplification of histone genes by circular chromosome formation in *Saccharomyces cerevisiae*. *Nature* 2006;**443**:1003–7.
5. Cohen S, Houben A, Segal D. Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. *Plant J* 2008;**53**:1027–34.
6. Verhaak RGW, Bafna V, Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer* 2019;**19**:283–8.
7. Wu S, Turner KM, Nguyen N, *et al.* Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* 2019;**575**:699–703.
8. Hung KL, Yost KE, Xie L, *et al.* ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature* 2021;**600**:731–6.
9. Yi E, Gujar AD, Guthrie M, *et al.* Live-cell imaging shows uneven segregation of extrachromosomal DNA elements and transcriptionally active extrachromosomal DNA hubs in cancer. *Cancer Discov* 2022;**12**:468–83.
10. Hu J, Zhang Z, Xiao S, *et al.* Microhomology-mediated circular DNA formation from oligonucleosomal fragments during spermatogenesis. *Elife* 2023;**12**:RP87115. https://doi.org/10.7554/eLife.87115.
11. Wang Y, Wang M, Djekidel MN, *et al.* eccDNAs are apoptotic products with high innate immunostimulatory activity. *Nature* 2021;**599**:308–14.
12. 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**:891–7.
13. Moller HD, Parsons L, Jorgensen TS, *et al.* Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci U S A* 2015;**112**:E3114–22.
14. Shoura MJ, Gabdank I, Hansen L, *et al.* Intricate and cell type-specific populations of endogenous circular DNA (eccDNA) in *Caenorhabditis elegans* and *Homo sapiens*. *G3 (Bethesda)* 2017;**7**: 3295–303.
15. Deshpande V, Luebeck J, Nguyen NPD, *et al.* Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat Commun* 2019;**10**:392.

16. Prada-Luengo I, Krogh A, Maretty L, Regenberg B. Sensitive detection of circular DNAs at single-nucleotide resolution using guided realignment of partially aligned reads. *BMC Bioinformatics* 2019;**20**:663.
17. Shibata Y, Kumar P, Layer R, *et al.* Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* 2012;**336**:82–6.
18. Zhang P, Peng H, Llauro C, *et al.* ecc_finder: a robust and accurate tool for detecting extrachromosomal circular DNA from sequencing data. *Front Plant Sci* 2021;**12**:743742. https://doi.org/10.3389/fpls.2021.743742.
19. Mann L, Seibt KM, Weber B, Heitkam T. ECCsplorer: a pipeline to detect extrachromosomal circular DNA (eccDNA) from next-generation sequencing data. *BMC Bioinformatics* 2022;**23**: 40.
20. Henriksen RA, Jenjaroenpun P, Sjøstrøm IB, *et al.* Circular DNA in the human germline and its association with recombination. *Mol Cell* 2022;**82**:209–17 e207.
21. Wanchai V, Jenjaroenpun P, Leangapichart T, *et al.* CReSIL: accurate identification of extrachromosomal circular DNA from long-read sequences. *Brief Bioinform* 2022;**23**:bbac422. https://doi.org/10.1093/bib/bbac422.
22. Koche RP, Rodriguez-Fos E, Helmsauer K, *et al.* Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat Genet* 2020;**52**:29–34.
23. Zhao XK, Xing P, Song X, *et al.* Focal amplifications are associated with chromothripsis events and diverse prognoses in gastric cardia adenocarcinoma. *Nat Commun* 2021;**12**:6489.
24. Hung KL, Luebeck J, Dehkordi SR, *et al.* Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH. *Nat Genet* 2022;**54**:1746–54.
25. Ewels PA, Peltzer A, Fillinger S, *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* 2020;**38**:276–8.
26. Gao X, Liu K, Luo S, *et al.* Comparative analysis of methodologies for detecting extrachromosomal circular DNA. *bioRxiv* 2023: 2023.12.01.569546. https://www.biorxiv.org/content/10.1101/2023.12.01.569546.
27. Krueger F. Trim Galore!: a wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. *Babraham Institute* 2015.
28. Heinz S, Benner C, Spann N, *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;**38**:576–89.
29. Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Babraham Institute, Cambridge, UK: Babraham Bioinformatics, 2010.
30. Ritchie ME, Phipson B, Wu D, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47. https://doi.org/10.1093/nar/gkv007.
31. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40.
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.
33. Gu Z, Gu L, Eils R, *et al.* Circlize implements and enhances circular visualization in R. *Bioinformatics* 2014;**30**:2811–2.
34. Wu T, Hu E, Xu S, *et al.* clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021;**2**:100141.
35. Koster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 2012;**28**:2520–2.
36. Zhong T, Wang W, Liu H, *et al.* eccDNA Atlas: a comprehensive resource of eccDNA catalog. *Brief Bioinform* 2023;**24**:bbad037. https://doi.org/10.1093/bib/bbad037.
37. Aslam FA, Mohammed HN, Lokhande PS. Efficient way of web development using python and flask. *Int J Adv Res Comput Sci* 2015.
38. 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**:1069.
39. Zhu Y, Liu Z, Guo Y, *et al.* Whole-genome sequencing of extrachromosomal circular DNA of cerebrospinal fluid of medulloblastoma. *Front Oncol* 2022;**12**:934159. https://doi.org/10.3389/fonc.2022.934159.
40. Helmsauer K, Valieva ME, Ali S, *et al.* Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. *Nat Commun* 2020;**11**:5823.
41. Li F, Ming W, Lu W, *et al.* FLED: a full-length eccDNA detector for long-reads sequencing data. *Brief Bioinform* 2023;**24**:bbad388. https://doi.org/10.1093/bib/bbad388.
42. Giurgiu M, Wittstruck N, Rodriguez-Fos E, *et al.* Decoil: reconstructing extrachromosomal DNA structural heterogeneity from long-read sequencing data. *bioRxiv* 2023:2023.11.15.567169. https://www.biorxiv.org/content/10.1101/2023.11.15.567169.
43. Jiang R, Yang M, Zhang S, Huang M. Advances in sequencing-based studies of microDNA and ecDNA: databases, identification methods, and integration with single-cell analysis. *Comput Struct Biotechnol J* 2023;**21**:3073–80.