

RESEARCH

Open Access



# Sequencing of methylase-accessible regions in integral circular extrachromosomal DNA reveals differences in chromatin structure

Weitian Chen<sup>1,2†</sup>, Zhe Weng<sup>2†</sup>, Zhe Xie<sup>2,3</sup>, Yeming Xie<sup>2</sup>, Chen Zhang<sup>2</sup>, Zhichao Chen<sup>1,2</sup>, Fengying Ruan<sup>2</sup>, Juan Wang<sup>2</sup>, Yuxin Sun<sup>2</sup>, Yitong Fang<sup>2</sup>, Mei Guo<sup>2</sup>, Yiqin Tong<sup>2</sup>, Yanning Li<sup>2</sup> and Chong Tang<sup>2,4,5\*</sup> 

## Abstract

**Background:** Although extrachromosomal DNA (ecDNA) has been intensively studied for several decades, the mechanisms underlying its tumorigenic effects have been revealed only recently. In most conventional sequencing studies, the high-throughput short-read sequencing largely ignores the epigenetic status of most ecDNA regions except for the junctional areas.

**Methods:** Here, we developed a method of sequencing enzyme-accessible chromatin in circular DNA (CCDA-seq) based on the use of methylase to label open chromatin without fragmentation and exonuclease to enrich ecDNA sequencing depth, followed by long-read nanopore sequencing.

**Results:** Using CCDA-seq, we observed significantly different patterns in nucleosome/regulator binding to ecDNA at a single-molecule resolution.

**Conclusions:** These results deepen the understanding of ecDNA regulatory mechanisms.

**Keywords:** ecDNA, Chromatin accessibility, Methylation, m<sup>6</sup>A, Methyltransferase

## Introduction

Finding a cure for cancer has been a challenge for various reasons, such as oncogene amplification, tumor evolution, and genetic heterogeneity [9, 31, 36]. Recently, it has been demonstrated that circular extrachromosomal DNA (ecDNA) plays a critical role in carcinogenesis, as it promotes oncogene amplification [40], drives tumor evolution, and contributes to genetic heterogeneity [28, 37, 38]. Circular ecDNA is arranged next to chromatin in a circular structure, featuring the head-to-tail junctional sequence and distal homologous genome sequence. The cancer-specific ecDNA may have an average size of 1.3 MB [4]. Although ecDNA has been known since 1964

[28], the elucidation of its role has been slow due to the lack of adequate molecular analytical techniques [2].

The development of new techniques, including computational advances, enabled genetic and epigenetic studies of ecDNA, and attempts have been made to identify ecDNA from sequencing data using improved algorithms from specificity and sensitivity. Most algorithms, such as Circle-Map [30], AmpliconArchitect [7], and CIRC\_finder [21], relied on the detection of the ecDNA junction sequence and enabled ecDNA identification in numerous cancer tissues [18, 21, 29, 37], aging cells [15], blood plasma [20, 41], and healthy somatic tissues [26]. However, due to the rareness of ecDNA in sequencing data, these approaches require the enrichment of ecDNA molecules; for example, circular DNA is obtained by the digestion of linear DNA with nucleases, followed by rolling circle amplification [25]. To further improve the

\*Correspondence: leochong718@gmail.com

<sup>†</sup>Weitian Chen and Zhe Weng have contributed equally to this work

<sup>2</sup> BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China

Full list of author information is available at the end of the article



accuracy of ecDNA detection, the long-read sequencing technology has been used to verify the ecDNA junction structure [6, 24]. However, functional epigenetic studies of ecDNA are currently lacking. Given the increasing awareness of ecDNA and its role in oncogene expression, understanding genome-wide ecDNA chromatin state and transcription status is essential. The most recent and advanced theory proposed by Wu et al. offers insights into highly accessible chromatin region and high expression of oncogenes located within these regions in ecDNA using the assay for transposase-accessible chromatin sequencing (ATAC-seq), and chromatin immunoprecipitation sequencing (ChIP-seq) [40]. Moreover, few studies have examined genome-wide ecDNA epigenome, because it is difficult to analyze ecDNA junction structure and epigenome information simultaneously.

Moreover, ATAC-seq and ChIP-seq analyses are based on the peak calling algorithm of populated fragments, and they do not consider the molecular heterogeneity. Therefore, there are benefits for expanding sequencing length and achieving single molecular resolution of the ecDNA epigenome owing to the limitations due to the short-read sequencing and short fragmentation required for ATAC-seq [3] and MNase-seq [32]. Our study was prompted by the existing long-read sequencing methods for assessing chromatin state, such as MeSMLR-seq [39], SAMOSA [1], nanoNOMe-seq [22], SMAC-seq [33], and fiber-seq [35]. We used the N6-methyladenosine (m<sup>6</sup>A) methyltransferase EcoGII to soft label accessible chromatin regions without fragmentation and named this method “sequencing of enzyme-accessible chromatin in circular DNA” (CCDA-seq). Using this method, we enriched ecDNA by digesting the linear genome using nuclease. Nanopore sequencing accurately detected the m<sup>6</sup>A-probed ecDNA regions of accessible chromatin and junctional structure properties simultaneously in the long range. Using CCDA-seq, we found a high diversity of ecDNA regions of accessible chromatin and their coordination with distal regulators at a single-molecule resolution, which has not been reported before.

## Results

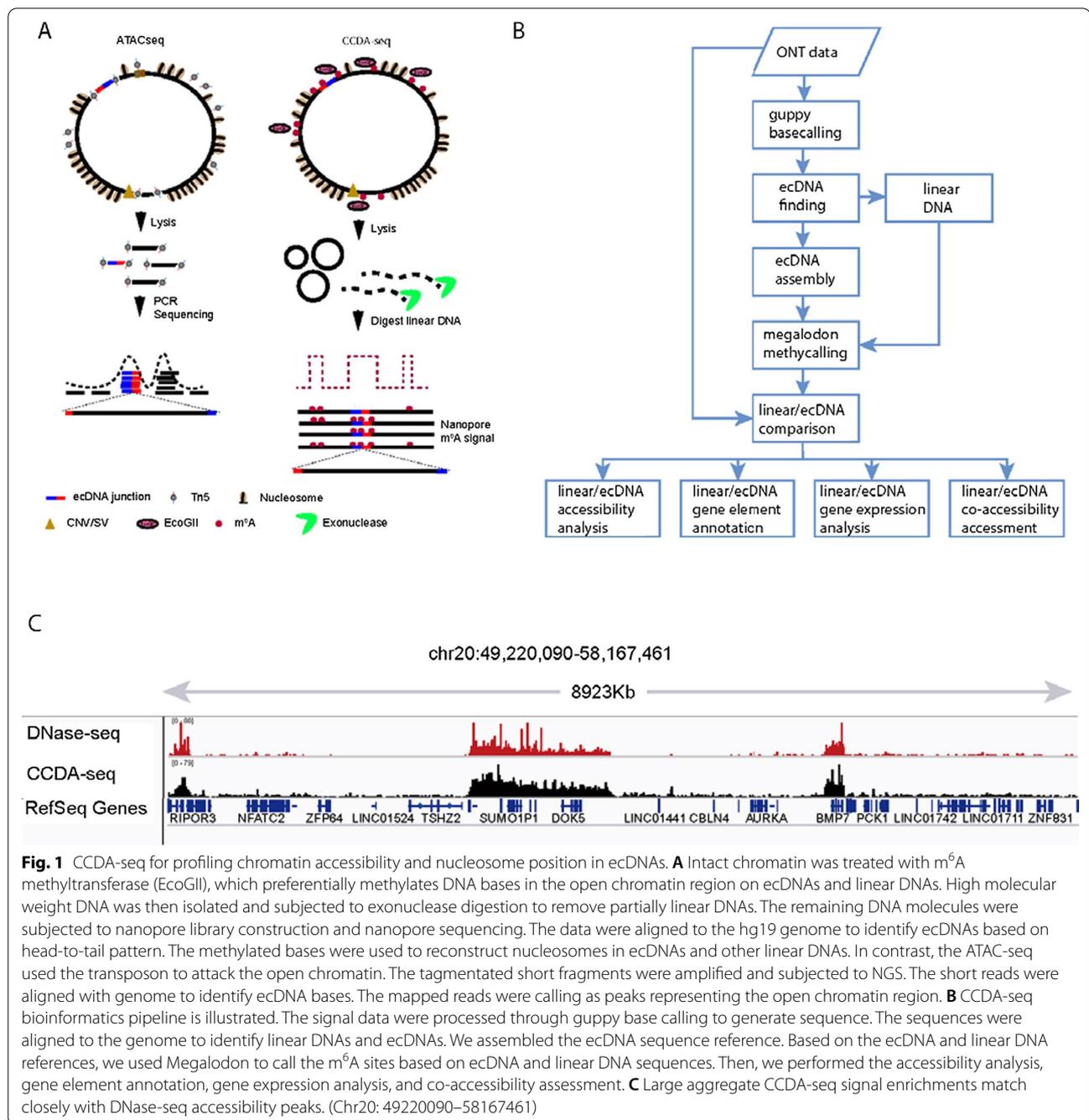
### CCDA-seq comprehensively maps accessible chromatin in ecDNA at a multikilobase scale

ecDNA plays an important role in tumorigenesis due to the high accessibility of its chromatin and carried oncogenes [40]. Conventional approaches to study chromatin accessibility are based on the concept that chromatin protects the bound sequence from attack by transposase (Fig. 1A) or MNase [32]. In ATAC-seq, the open, accessible genome region is first preferentially tagged using transposase, followed by next-generation sequencing (NGS) (Fig. 1A). However, this method is not employed

in most integral ecDNA chromatin studies due to the homologous ecDNA/genome sequences, making the distinction between ecDNA and linear genome DNA difficult. In general, previous studies of ecDNA chromatin based on NGS of short reads only observed the chromatin status in the junction region (200 bp around the junction) and bioinformatically analyzed other distal ecDNA areas because of limitations of the techniques used (>200 bp to junction regions) (Fig. 1A). To solve these problems, we built a generalized framework based on the concept of the MeSMLR-seq (Wang et al. 2019), SAMOSA (Abdulhay et al. 2020), SMAC-seq [33], and fiber-seq [35]. We applied soft labeling with the m<sup>6</sup>A methyltransferase EcoGII that preferentially methylates the adenosine in the openly accessible DNA region without fragmentation by a transposase (Fig. 1A). To improve the ecDNA capturing efficiency, the exonuclease was introduced to remove the linear genome DNA [8]. The integral ecDNA was sequenced by nanopore sequencing, and the probed m<sup>6</sup>A was detected [33]. By analysing the generated data, we first identified ecDNA molecules by head-to-tail junction locations and by dynamically mapping the segments of sequences to the genome (Fig. 1B). Based on the head-to-tail junction locations, we then reassembled the partial ecDNA sequences as the new reference and identified the m<sup>6</sup>A signal based on the reassembled ecDNA sequence to prevent signal bias in the junction region (Fig. 1B).

The statistical analysis showed that the read length was between 10 and 100 kb, which was 50× broader than the junctional region observed in conventional ATAC-seq (Additional file 1: Figure S1). The long-read feature also makes the nanopore sequencing method optimal for applications, such as structure variation (SV), copy number variation (CNV), and ecDNA identification with better sensitivity and specificity [14]. As expected, 80% of ecDNA molecules detected in our CCDA-seq could be validated through PCR (Additional file 1: Figure S2). ecDNA and residual linear DNA accounted, respectively, for 0.9% and 99.1% of the total sequencing reads (Additional file 1: Figure S3) after exonuclease treatment, in which ecDNA was enriched ninefold (Additional file 1: Figure S4). To further explore ecDNA functionality, we assembled ecDNAs to reconstruct the full length of ecDNA, the length of which ranged from 2 kb to 1 MB (Additional file 1: Figure S5). Seventy-five percent of ecDNA contained genes (Additional file 1: Figures S6, S7), suggesting the translational function of ecDNA.

The m<sup>6</sup>A probability distribution in *Megalodon* showed two distinct peaks for the treated sample. The distribution of the narrow peak with lower m<sup>6</sup>A probability (mean=0.49) was similar to the background noise distribution (Additional file 1: Figure S8). Therefore, we set



m<sup>6</sup>A methylation probability over 0.53 as the cutoff for the true m<sup>6</sup>A signal (Additional file 1: Figure S8). The real positive cutoff value was set at 0.53, and the m<sup>6</sup>A calling specificity and sensitivity were 0.99 and 0.92, respectively (Additional file 1: Figure S4). The residual linear DNA was used as internal control for validation using published ATAC-seq data [11]. CCDA-seq achieved consistency and coherence with ATAC-seq data in various resolutions (Fig. 1C, Additional file 1: Figures S9,

S10). The m<sup>6</sup>A labeling deviation was inversely proportional to the m<sup>6</sup>A ratio and strongly reduced to 0.015 in m<sup>6</sup>A-enriched region (Additional file 1: Figure S11). The impact of the exonuclease treatment and reproducibility have been also investigated (Additional file 1: Figure S12). These characteristics of CCDA-seq were critical for effectively measuring the accessibility of chromatin in linear and circular DNA molecules in the multikilobase range.

Another remarkable feature of CCDA-seq is that it enables investigation of the ecDNA chromatin status at a single-molecule resolution, at which the single base m<sup>6</sup>A probability varied from 0.6 to 1. (Additional file 1: Figure S13). In practice, the resolution of accessible chromatin regions was around 200 bp. We adopted a Bayesian procedure to aggregate methylation probabilities and derived the accurate single-molecule accessibility calls over windows of arbitrary size (Additional file 1: Figure S13). In summary, CCDA-seq offers attractive features in terms of the elucidation of the integral ecDNA chromatin status in the multikilobase range at a single-molecule resolution.

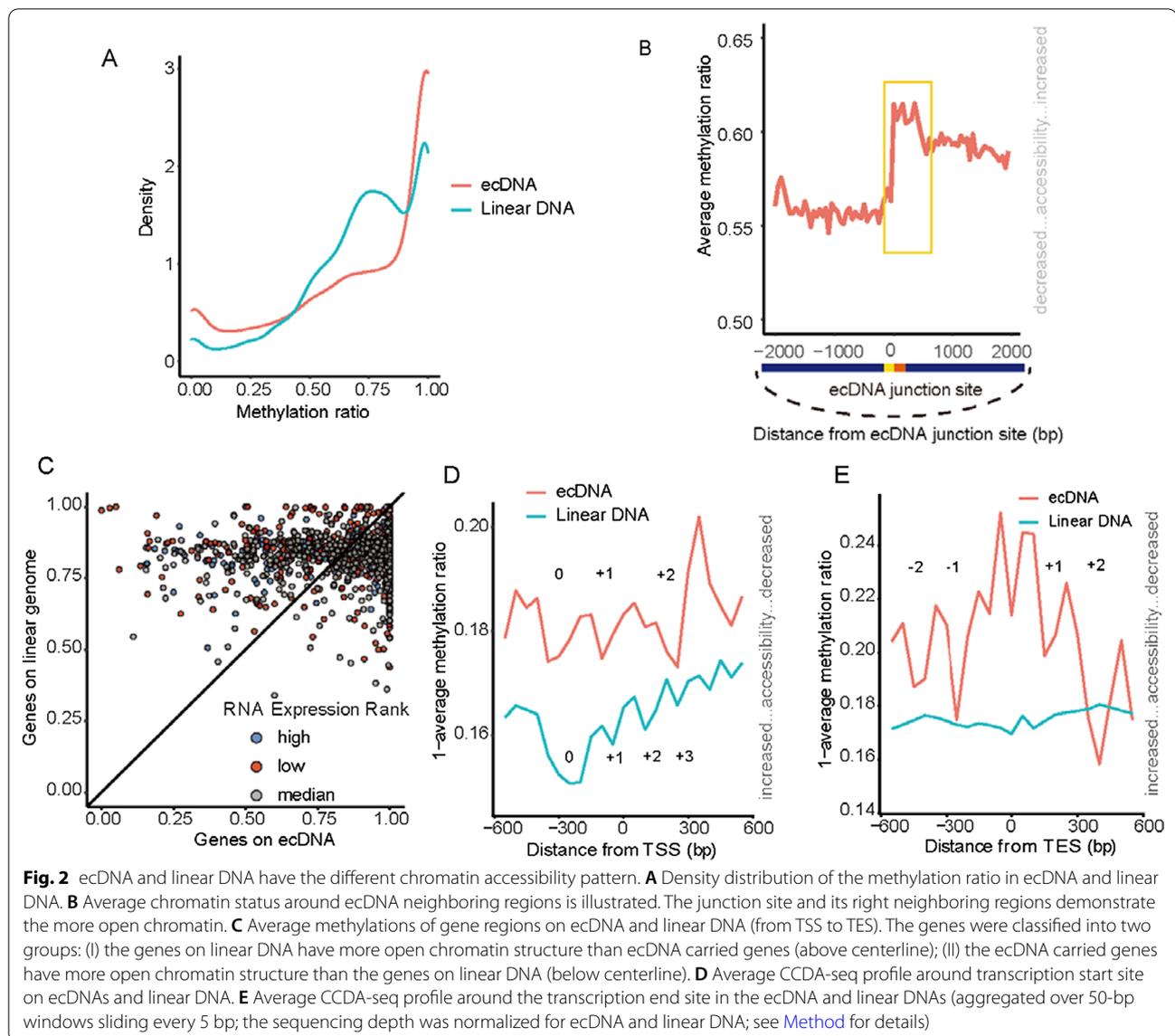
#### Diverse patterns of ecDNA chromatin accessibility

Evidence from other studies that utilized ATAC-seq and Chip-seq suggests that the active chromatin status and highly accessible ecDNA chromatin may be associated with high levels of oncogene transcription [40]. To distinguish ecDNA molecules from linear DNA molecules in ATAC-seq and Chip-seq, it is necessary to screen out the short reads (~200 bp) spanning the non-homologous end-joining ecDNA sequence. One problem with these approaches is the potential bias to neglect the distal regions due to focusing on the ~200 bp reads neighboring ecDNA junctional sequences. CCDA-seq, as a long-read technology, may facilitate precise ecDNA detection [6, 14, 26] and observation of the distal chromatin status in integral ecDNA. We obtained an extensive catalog of 12,997 different ecDNA molecules formed from chromosomal breakpoints between 0.05 and 100 kb (Additional file 2: Table S1). Gene ontology (GO) analysis of the genes harbored by these ecDNA molecules revealed significant enrichment in the GO terms GTPase-related activity, channel activity, and nucleoside-triphosphatase activity, i.e., processes playing essential roles in cancer progression (Additional file 1: Figure S14) [13, 16]. RNA-seq data analysis showed that there were 340 highly expressed ecDNA genes (25% rank), 464 moderately expressed genes (25–75% rank), and 589 genes with low expression (75–100% rank), indicating that not all ecDNA genes are highly expressed.

By comparing the average chromatin accessibility between ecDNA and homologous linear DNA, we found that the ecDNA chromatin is twofold more accessible than that the linear DNA chromatin (Fig. 2A). These findings reinforce the general notion that ecDNA amplification results in higher oncogene transcription [40], coupled with the enhanced chromatin accessibility in the junctional region. The CCDA-seq data were subjected to the detailed mapping of the ecDNA chromatin status. We found that chromatin in the ecDNA junctional areas is significantly more accessible than in

other junctional distal regions (Fig. 2B). This is an interesting finding, as it suggests that the conclusions drawn by observing only the junctional areas after the conventional ATAC-seq may be biased and not necessarily relate to the whole ecDNA chromatin. We calculated average fractions of m<sup>6</sup>A methylation from the gene transcription start site (TSS) to the gene transcription end site (TES) on each gene-spanning read. A pairwise scatter plot of the average accessibility between ecDNA genes and linear genome genes showed that 62% of gene regions are more accessible in the ecDNA than in the linear DNA (Fig. 2C) (considering that standard deviation is 5%). A comparison of ecDNA and linear DNA chromatin profiles around the TSS/TES (+/- 500 bp) revealed a significant difference in nucleosome depletion/occupancy patterns (Fig. 2D, E). The nucleosome organization may impact access to ecDNA (Fig. 2D, E). Considering that 62% of gene regions were more accessible in the ecDNA than in the linear DNA, we further plotted the chromatin structure around TSS/TES (+/- 500 bp) of these genes (Additional file 1: Figure S15). The formation of nucleosome depletion regions (NDRs) on linear DNA is restricted to 200 bp before TSSs. In contrast, the NDRs in ecDNA are distributed uniformly (Additional file 1: Figure S15). The other 36% of gene regions are more accessible in the linear DNA than in the ecDNA. The TSSs/TEs (+/- 500 bp) were also significantly more accessible in the linear DNA than in the ecDNA with different NDR patterns (Additional file 1: Figure S16). The formation of large NDRs was restricted to TSSs in the linear DNA, which was not observed in the ecDNA.

Another illustration of the complex interplay between chromatin states in the ecDNA and linear DNA relates to the transcriptional activity. The chromatin of the linear DNA active genes (top 25% rank) is largely devoid of nucleosomes on TSSs due to the extremely high transcription activity (Additional file 1: Figure S17). In contrast, the chromatin structure of the ecDNA active genes adopts a distinct conformation, implying that ecDNA is regulated by different mechanisms (Additional file 1: Figure S17). For the transcriptionally inactive genes, the stationary nucleosome states are shown in the linear DNAs (Additional file 1: Figure S18). In contrast, ecDNA molecules still have the active nucleosome organization in the regions of 300 bp before TSSs, suggesting that chromatin accessibility is necessary but not sufficient for the enhancer or promoter activity in the ecDNA (Additional file 1: Figure S18). In conclusion, ecDNA and linear DNA have significantly different nucleosome depletion/occupancy patterns in various conditions, suggesting their distinct gene regulatory mechanisms.

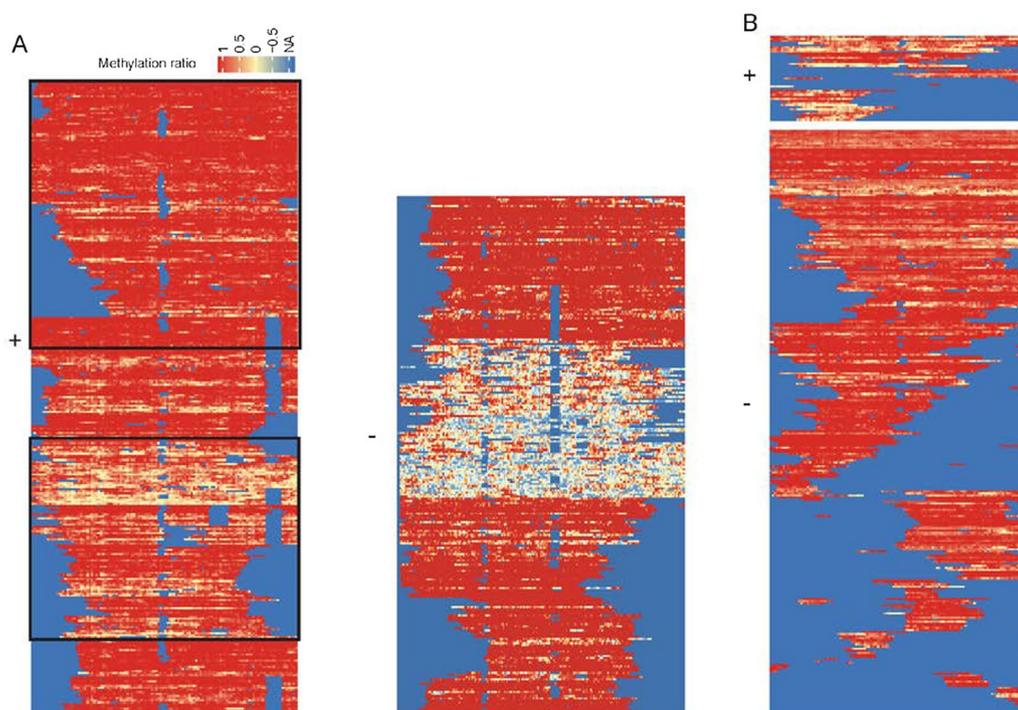


### Chromatin status in the ecDNA and linear genome DNA at a single-molecule resolution

The conventional ATAC-seq is based on statically calling the peak of the enriched read in a specific region [3]. Recent single-molecule and single-cell accessibility measurements suggested that ATAC-seq of cell populations represents an ensemble average of distinct molecular states [17]. An essential attribute of the CCDA-seq is a possibility to determine ecDNA chromatin accessibility at a single-molecule resolution by taking the advantage of small variance (Additional file 1: Figure S11) and increased cumulative probability in segments (Additional file 1: Figure S13). Measuring chromatin accessibility of

the single linear DNA has also been done using SMAC-seq [33] and fiber-seq [35].

We then asked whether CCDA-seq could reveal multiple chromatin accessibility states in ecDNA. The chromatin structure of the linear DNA (chr10: 42383201–42389251) adopts two distinct conformations: an inactive nucleosomal state and a state largely devoid of nucleosomes due to extremely high transcription activity [5] (Fig. 3A). It is thought that ecDNA in the majority of cancer cells exhibits active nucleosome status. As expected, 70% of ecDNA molecules come from the very active chromatin state (Fig. 3A). We observed highly heterogeneous nucleosome depletion/occupancy patterns



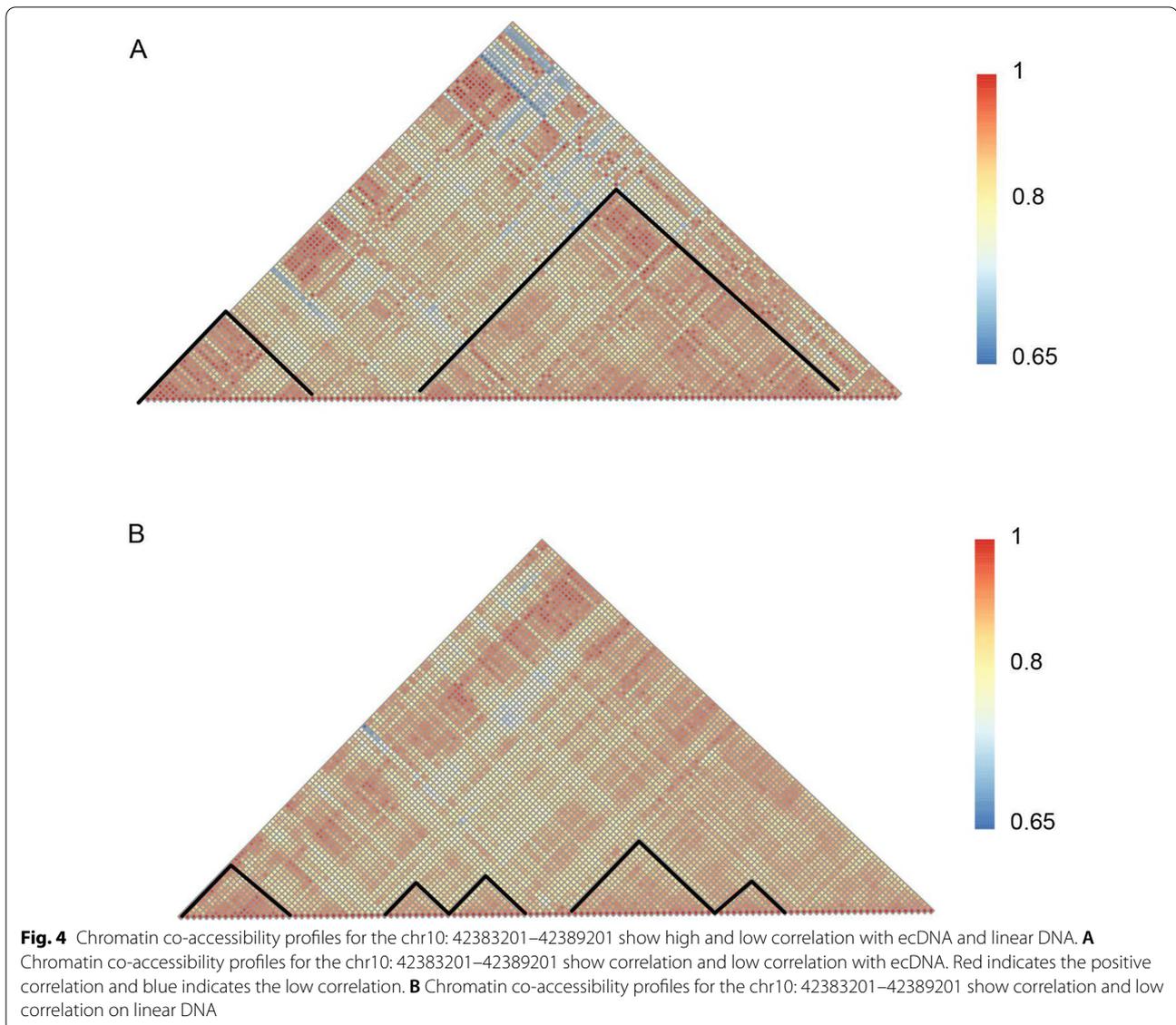
**Fig. 3** CCDA-seq reveals the distribution of alternative chromatin states in ecDNA arrays. **A** Shown are all reads covering the linear DNA region chr10: 42383201–42389201. The box highlights the active and inactive chromatin. **B** Shown are all reads covering the ecDNA region chr10: 42383201–42389201. The upper panel indicates the positive strand, and the lower panel indicates the negative strand

in ecDNA, and most chromatin molecules were not very active in the positive strand, suggesting distinct transcriptional regulation of ecDNA (Fig. 3B, upper panel). Some regulator enzymes may occupy the positive strand and restrict chromatin accessibility. The highly active ecDNA chromatin was also observed in other regions (Additional file 1: Figure S19). To avoid a conclusion biased by the heterogeneous activity of the methylase, other upstream and downstream regions were chosen as quality controls.

To further explore CCDA-seq resolution limits, we studied methylation patterns in more detail. In particular, we quantified strand-specific DNA accessibility and observed a strand-asymmetric DNA accessibility pattern in the linear genome (Additional file 1: Figure S19). The strand-asymmetric DNA accessibility pattern was also observed in ecDNA, and both strands displayed high heterogeneity (Fig. 3B, Additional file 1: Figure S19). This strand-specific heterogeneity in methylation potential within the nucleosome may inform about how transcription factors interact with nucleosome-associated DNA *in vivo*.

Wu et al. showed that ecDNA enables ultra-long range chromatin contact, permitting distant interactions

with regulatory elements [40]. We next examined co-accessibility patterns in the ecDNA and linear genome DNA by assessing nucleosome positioning correlations. The nucleosomes have higher correlation values in the ecDNA than in the linear DNA (Fig. 4A, B, Additional file 1: Figure S20). Moreover, ecDNA and linear DNA adopt significantly different chromatin co-accessibility patterns (Fig. 4A, B; Additional file 1: Figure S20). Average co-accessibility profiles in the linear DNA revealed a detectable correlation between nucleosome positions up to two to three nucleosomes away. For the ecDNA, this correlation was further and up to 20 nucleosomes away (Fig. 4A, B; Additional file 1: Figure S20). These results agree with the high-resolution chromosome conformation capture (HiC) result [40] in that the ecDNA is characterized by the distant chromatin interaction. It was interesting to note that ecDNA demonstrated some ultra-distant anticorrelated states. Overall, ecDNA molecules were highly heterogeneous and exhibited remote chromatin interactions, suggesting their different regulation mechanisms compared to those acting on linear DNAs.



## Discussion

Understanding of ecDNA functions may prove to be essential for the elucidation of tumorigenesis mechanisms [28, 37, 38]. Many ecDNA molecules have been identified in various cancer tissues [18, 21, 29, 37]. There has been an increasing research focus on the status of ecDNA chromatin to resolve the problem of ecDNA oncogene amplification [40]. However, most studies focused on short sequencing reads with junctional sequences detected to avoid the false-positive identification of ecDNA and to precisely determine the ecDNA epigenetic status. A large subgroup (60%) of ecDNAs covered regions that are not unique in the reference genome, which complicated their identification [27]. In this study, we used nanopore sequencing

to evaluate integral ecDNA chromatin accessibility in ecDNA long strands by applying m<sup>6</sup>A methyltransferase to label open chromatin without fragmentation. Consistent with the previously reported findings [40], most ecDNA genomic regions are more accessible than the linear genomic regions (Fig. 2A). For genetic regions, 63% of ecDNA molecules carried genes with more accessible chromatin structure than that of the linear DNA. However, in the remaining fraction of ecDNA (37%), chromatin of the gene regions was less accessible than in the corresponding linear DNA parts. Notably, the nucleosome depletion/occupancy patterns were significantly different between ecDNA and linear DNA. Our single-molecule resolution method allows footprinting of protein and nucleosome binding as well

as determination of the epigenetic signature of chromatin accessibility. It is hoped that this study will contribute to more comprehensive understanding of the ecDNA epigenome regulation.

In our experiments, we treated DNA samples with an exonuclease that removed the majority of the linear DNA molecules and increased the sequencing depth for the ecDNA (0.9%). Some identified linear DNA molecules may be generated from the ecDNA homologous regions without junctions, but the likelihood of that was around 0.9%, which is negligible. Compared with the parameters in the non-digest direct sequencing, we only obtained 0.1% of ecDNA-related reads (Additional file 1: Figure S4). The circular ecDNA enrichment was ninefold. The exonuclease treatment not only improved ecDNA sequencing coverage, but also ecDNA detection specificity (Additional file 1: Figure S2). However, DNA purification process could damage large-size ecDNA molecules over 1 Mb [34]. Such damaged ecDNA could be digested during exonuclease digestion and missed in the sequencing. A method that gently purifies large DNA molecules would be preferable in further large-scale ecDNA studies.

Megalodon is the latest software (compared with Tombo) that was chosen for m<sup>6</sup>A signal calling. In the ecDNA m<sup>6</sup>A calling, Tombo ignored half of the sequences or lost most ecDNA molecules for unknown reasons (Additional file 1: Figure S21). The sensitivity of Tombo for ecDNA m<sup>6</sup>A signals was 83% less than that of Megalodon. Although Megalodon improved the sensitivity of ecDNA m<sup>6</sup>A calling, it did not address the issue of the false-positive m<sup>6</sup>A signal, that most adenosine bases could be recognized as m<sup>6</sup>A with a probability of 0.4–1 using Megalodon. The only known way to solve the false positive issue is to employ data training with negative control samples (Additional file 1: Figure S8). We used 0.53 as m<sup>6</sup>A probability cutoff, successfully discriminating the m<sup>6</sup>A and false-positive signals with sensitivity of 0.92 and specificity of 0.99. In general, Megalodon performed better in ecDNA analysis, and its specificity improved following data training.

In the sequencing data, we found that the methylated treated DNA generated more data than the non-methylated DNAs, which was not consistent with the SMAC-seq and fiber-seq data [33, 35]. The highly open chromatin with highly methylated sites may have been enriched using our method. In our laboratory experiments, we found that the heavily modified DNA was more resistant to exonuclease digestion and, therefore, enriched as a result of enzymatic treatment. The non-treated sample showed a lower overall methylation level (Additional file 1: Figure S22). However, the nucleosome occupancy positions were not significantly affected by the exonuclease treatment (Additional file 1: Figure

S23). Moreover, in the strand-specific view, the reverse strand reads are generally less abundant than the positive strands. This may also be due to the different patterns of methylation of the positive and negative strands, which could result in different digestion efficiencies. This problem is usually overcome by increasing sequencing depth and using normalization methods. We also suggest that sequencing both treated and non-treated samples for ecDNA sequencing coverage will further improve quantification accuracy.

Only 63% of gene regions were within highly accessible chromatin in our experiments. However, Wu et al. showed using ATACseq technology that ecDNA molecules are mostly located in highly accessible chromatin [40]. When comparing results from all other regions with the published data, a good agreement was found in that 80% of ecDNA areas were highly accessible (Additional file 1: Figure S24). Most of the areas of highly accessible chromatin were distributed in the intron and intergenic regions (Additional file 1: Figure S25). The reasons for this remain unclear, but our results indicate that ecDNA has a highly open chromatin structure, especially in the intergenic and intronic regions.

## Conclusion

CCDA-seq is useful for studying the chromatin status of integral ecDNA, offering deep insights into the distinct mechanisms of ecDNA regulation. However, the ecDNA enrichment step requires exonuclease treatment, causing the loss of mega ecDNAs. It is assumed that future advances will help address the problems of DNA damage during the purification and the insufficient sequencing depth. The CCDA-seq will help the scientific community to understand different mechanisms of ecDNA regulation, especially in cancer development.

## Method

### Cell culture

Human mammary gland carcinoma cell line MCF-7 was obtained from ATCC. MCF-7 were grown in DMEM (Gibco 11995065) supplemented with 10% FBS (Gibco, 10099141), 0.01 mg/ml insulin (MedChemExpress, HY-P1156), and 1% penicillin–streptomycin (Gibco, 15140122). The cell line was regularly checked for mycoplasma infection (Yeasen, 40612ES25).

### Nuclei isolation and MTase treatment

Cells were grown to 70–80% confluency, and were collected by TrypLE (Gibco, 12604013). After 300×g centrifuge for 5 min, nuclei were isolated with lysis buffer (100 mM Tris–HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>,

0.1 mM EDTA, 0.5% CA630) for 5 min on ice. Nuclei were centrifuged at  $300\times g$  in wash buffer (100 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1 mM EDTA) at 4 degree, and washed twice for 5 min and counted.

$1 \times 10^6$  intact nuclei were subjected to an m<sup>6</sup>A methylation reaction mixture containing  $1 \times$  Cutsmart buffer (NEB), 200U of non-specific adenine methyltransferase M.EcoGII (NEB, M0603S), 300 mM sucrose, and 96  $\mu$ M S-adenosylmethionine (NEB, B9003S) in 500  $\mu$ l volume. The reaction mixture was set up at a 37-degree thermomixer with shaking at 1000 rpm for 30 min. S-adenosylmethionine was replenished at 640  $\mu$ M every 7.5 min at 7.5, 15, and 22.5 min into the reaction mixture. The reaction was stopped by adding an equal volume of stop buffer (20 mM Tris-HCl pH 7.4, 600 mM NaCl, 1% SDS, 10 mM EDTA). No methylation controls were treated in the same conditions without adding M.EcoGII in the reaction mixture. The samples were then treated with 20  $\mu$ l of Proteinase K (20 mg/ml) at 55 degrees overnight, and the DNA was extracted with phenol: chloroform extraction and ethanol precipitation.

#### ecDNA isolation, purification, and sequencing

ecDNA was isolated by Circle-Seq [25] method, which digested linear DNA with modifications. Briefly, 10  $\mu$ g of M.EcoGII treated DNA was subjected to a reaction mixture containing  $1 \times$  plasmid-safe reaction buffer, 20U plasmid-safe ATP-dependent DNase (Lucigen, E3101K), 1 mM ATP, and nuclease-free water was supplemented to a final volume of 100  $\mu$ l. The reaction mixture was incubated at 37 degrees for 7 days. Every 24 h, the reaction mixture was replenished by adding 20U plasmid-safe ATP-dependent DNase, 1 mM ATP, and 0.4  $\mu$ l 10X plasmid-safe reaction buffer. Digested ecDNA was purified with 1.8X AMPure XP beads (Beckman Coulter).

Purified ecDNA was prepared for nanopore sequencing by ligation kit LSK-SQK108(ONT). The samples were 10 kb by Covaris G tubes, end-repaired and dA-tailed using NEBnext Ultra II end-repair module (NEB), followed by clean-up using 1.8X AMPure XP beads. Sequencing adaptors and motor proteins were ligated to end-repaired DNA fragments using blunt/TA ligase master mix (NEB), followed by clean-up using 0.4  $\times$  AMPure XP beads. 1ug adaptor-ligated samples per flow cell were loaded onto PRO-002 flowcells and run on PromethION sequencers for up to 72 h. Data were collected by MinKNOW v.1.14.

#### Base-calling and linear DNA methylation calling

Reads from the ONT data were processed using Megalodon V2.2.9, which used Guppy base caller to base-calling, and Guppy model config `res_dna_r941_min_modbases-all-context_v001.cfg` released into the Rerio repository

was used to identify DNA m<sup>6</sup>A methylation. Megalodon\_extras was used to get per read modified\_bases from the Megalodon basecalls and mappings results. To further explore the accurate threshold of methylation probability, a control sample with almost no m<sup>6</sup>A methylation was used as background noise, and the Gaussian mixture model was used to fit the methylation probability distribution generated by Megalodon.

#### ecDNA calling

ONT Reads meet the following conditions were defined as ecDNA molecules performed by the inner mappy/minimap2 aligner [23]. (1) One segment (>1 kb) of an ONT read was mapped to the genome at one site, and another segment (>1 kb) was mapped to the genome at another site. (2) Two segments were mapped to the same chromosome. (3) Two segments were mapped to the same strand of the genome. (4) Two segments in a pair showed outward orientation.

#### Nanopore ecDNA methylation calling

Due to ecDNA special structure, the m<sup>6</sup>A calling cannot be successfully performed by aligning to the reference genome, especially for junctional regions. The custom python script was used to assemble ecDNA reference genome sequences according to the table generated from the previous step. Considering that the read length might be longer than the ecDNA reference, the ecDNA reference was subsequently preprocessed by adding 10 MN to the ends to increase the mapping efficiency. The downstream step is performed in a similar way as linear DNA methylation calling.

#### Circular DNAs assembly

Circular DNAs assembly was performed by Flye (v2.8.3-b1695) [19] assembler on the raw FASTQ files with the following settings: `-genome-size 1G -meta-nano-raw` [12]. Circular DNAs sequences were aligned to the hg19 genome by NCBI blastn(v2.2.28+). And we calculate the distribution and proportion of genes on Circular DNAs by custom script. Finally, the R packages circlize (v0.4.12.1004) [10] was used to visualize Circular DNAs.

#### Annotation and methylation configuration

TES, TTS, CDS, and other gene elements were downloaded from UCSC Table Browser. In addition, the gene elements were processed into 50 bp bin for downstream analysis. Linear DNA and ecDNA were also processed to the size of 50 bp bin and sliding for 5 bp. The accessibility score over multi base-pair windows was calculated as

methylation ratio =  $m^6A$  bases in all covered reads under bin/ adenosine bases in all covered reads under the bin.

### RNA-seq data analysis

The RNA-seq data of MCF-7 was downloaded from the Gene Expression Omnibus (GEO) repository database with the accession number GSE71862. The gene expression was divided into three categories: high, medium, and low, representing 25%, 25–75%, and 75% gene expression rank, respectively.

### Co-accessibility assessment

To evaluate co-accessibility patterns along the genome, we applied COA as follows. Each chromosome in the genome was split into windows of size  $w$ . For each such window  $(i, i + w)$ , we identified another window  $(j, j + w)$  such that the span  $(i, j, w)$  was covered by  $\geq N$  reads. For each single spanning molecule  $k$ , accessibility scores ( $A$ ) in each bin were then aggregated and binarized as described above. The local co-accessibility matrix between two windows was calculated:

$$COA_{i,j,w} = \text{mean}_N \left( 1 - \frac{|A_{i,w-A_j,w}|}{A_{i,w} + A_{j,w}} \right)$$

### Abbreviations

CCDA-seq: The sequencing of enzyme-accessible chromatin in circular DNA; ecDNA: Extrachromosomal DNA;  $m^6A$ : N<sup>6</sup>-methyladenosine; ATAC-seq: The assay for transposase-accessible chromatin sequencing; Chip-seq: Chromatin Immunoprecipitation Sequencing; nanoNOME-seq: Nanopore single-molecule Nucleosome Occupancy and Methylome sequencing; SMAC-seq: Single-molecule long-read accessible chromatin mapping sequencing assay; Mnase: Micrococcal nuclease; NGS: Next-generation sequencing; SV: Structure variation; CNV: Copy number variation; GO: Gene ontology; TSS: Transcription start site; TES: Gene transcription end site; NDRs: Nucleosome depletion regions; HiC: High-resolution chromosome conformation capture; ATAC-seq: Assay of transposase-accessible chromatin with visualization; CNGB: China National GeneBank.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13072-021-00416-5>.

**Additional file 1.** Additional Figures.

**Additional file 2: Table S1.** The profiles of the ecDNAs.

### Acknowledgements

Not applicable.

### Authors' contributions

CT designed and supervised the experiments. ZW and XZ perform the lab experiments; WTC performs the bioinformatics data analysis. All others joined the data analysis. All authors read and approved the final manuscript.

### Funding

This research was supported by the Science, Technology, and Innovation Commission of Shenzhen Municipality (Grant Number JSGG20170824152728492). The supporter had no role in designing the study, data collection, analysis, and interpretation, or in writing the manuscript.

### Availability of data and materials

Nanopore raw data are available at China National GeneBank (CNGB) with project number of CNP0001720.

### Declarations

#### Ethical approval and consent to participate

Not applicable.

#### Consent for publication

All subjects have written informed consent.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. <sup>2</sup>BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China. <sup>3</sup>Department of Biology, Cell Biology and Physiology, University of Copenhagen 13, 2100 Copenhagen, Denmark. <sup>4</sup>Nantong University, Nantong 226000, China. <sup>5</sup>Nephrosis Precision Medicine Innovation Center, University of Beihua School of Medicine, Jilin City 132011, China.

Received: 16 May 2021 Accepted: 11 August 2021

Published online: 23 August 2021

### References

- Abdulhay NJ, McNally CP, Hsieh LJ, et al. Massively multiplex single-molecule oligonucleosome footprinting. *Elife*. 2020;9:e59404.
- Bailey C, Shoura MJ, Mischel PS, Swanton C. Extrachromosomal DNA-relieving heredity constraints, accelerating tumour evolution. *Ann Oncol*. 2020;31:884–93.
- Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. 2015;109:21.
- Chiu RWK, Dutta A, Henssen AG, Lo YMD, Mischel P, Regenberg B. 2020. What is extrachromosomal circular DNA and what does it do? *Clin Chem*. 66: 754–759.
- Conconi A, Widmer RM, Koller T, Sogo J. Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell*. 1989;57:753–61.
- deCarvalho AC, Kim H, Poisson LM, Winn ME, Mueller C, Cherba D, Korman J, Seth S, Protopopov A, Felicella M, et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat Genet*. 2018;50:708–17.
- Deshpande V, Luebeck J, Nguyen N-PD, Bakhtiari M, Turner KM, Schwab R, Carter H, Mischel PS, Bafna V. 2019. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat Commun*. 10: 1–14.
- Gaubatz JW, Flores SC. Purification of eucaryotic extrachromosomal circular DNAs using exonuclease III. *Anal Biochem*. 1990;184:305–10.
- Gillies RJ, Verduzco D, Gatenby RA. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat Rev Cancer*. 2012;12:487–93.
- Gu Z, Gu L, Eils R, Schlesner M, Brors BJB. circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30: 2811–2812.
- He HH, Meyer CA, Chen MW, Jordan VC, Brown M, Liu XS. Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Res*. 2012;22:1015–25.

12. Helmsauer K, Valieva ME, Ali S, González RC, Schöpflin R, Röefzaad C, Bei Y, Garcia HD, Rodriguez-Fos E, Puiggròs MJNC. Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. *Nat Commun.* 2020;11:1–12.
13. House CD, Wang B-D, Ceniccola K, Williams R, Simaan M, Olender J, Patel V, Baptista-Hon DT, Annunziata CM, Silvio Gutkind J, et al. Voltage-gated Na<sup>+</sup> channel activity increases colon cancer transcriptional activity and invasion via persistent MAPK signaling. *Sci Rep.* 2015;5:11541.
14. Huddleston J, Chaisson MJ, Steinberg KM, Warren W, Hoekzema K, Gordon D, Graves-Lindsay TA, Munson KM, Kronenberg ZN, Vives L, et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* 2017;27:677–85.
15. Hull RM, King M, Piza G, Krueger F, Vergara X, Houseley J. Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLoS Biol.* 2019;17:e3000471.
16. Kazanietz MG, Caloca MJ. The Rac GTPase in cancer: from old concepts to new paradigms. *Can Res.* 2017;77:5445–51.
17. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet.* 2019;20:207–20.
18. Koche RP, Rodriguez-Fos E, Helmsauer K, Burkert M, MacArthur IC, Maag J, Chamorro R, Munoz-Perez N, Puiggròs M, Garcia HD. 2020. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat Genet.* 52: 29–34
19. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37:540–6.
20. 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;15:1197–205.
21. Kumar P, Kiran S, Saha S, Su Z, Paulsen T, Chatrath A, Shibata Y, Shibata E, Dutta A. 2020. ATAC-seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. *Sci Adv.* 6: eaba2489
22. Lee I, Razaghi R, Gilpatrick T, Molnar M, Gershman A, Sadowski N, Sedlazeck FJ, Hansen KD, Simpson JT, Timp W. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat Methods.* 2020;17:1191–9.
23. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34:3094–100.
24. Mehta D, Cornet L, Hirsch-Hoffmann M, Zaidi SS-A, Vanderschuren H. Full-length sequencing of circular DNA viruses and extrachromosomal circular DNA using CIDER-Seq. *Nat Protoc.* 2020;15:1673–89.
25. Møller HD. Circle-Seq: isolation and sequencing of chromosome-derived circular DNA elements in cells. *Methods Mol Biol.* 2020;2119:165–81.
26. Møller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P, Maretty L, Hansen AJ, Snyder MP, Pilegaard H, et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun.* 2018;9:1069.
27. Moller HD, Parsons L, Jorgensen TS, Botstein D, Regenberg B. Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci USA.* 2015;112:E3114–3122.
28. Paulsen T, Kumar P, Koseoglu MM, Dutta A. Discoveries of extrachromosomal circles of DNA in normal and tumor cells. *Trends Genet.* 2018;34:270–8.
29. Paulsen T, Shibata Y, Kumar P, Dillon L, Dutta A. 2019. Extrachromosomal circular DNA, microDNA, without canonical promoters produce short regulatory RNAs that suppress gene expression. *bioRxiv:* 535831.
30. 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 Bioinform.* 2019;20:663.
31. Ray Chaudhuri A, Callen E, Ding X, Gogola E, Duarte AA, Lee JE, Wong N, Lafarga V, Calvo JA, Panzarino NJ, et al. Replication fork stability confers chemoresistance in BRCA-deficient cells. *Nature.* 2016;535:382–7.
32. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. Dynamic regulation of nucleosome positioning in the human genome. *Cell.* 2008;132:887–98.
33. Shipony Z, Marinov GK, Swaffer MP, Sinnott-Armstrong NA, Skotheim JM, Kundaje A, Greenleaf WJ. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat Methods.* 2020;17:319–27.
34. Smith CL, Cantor CR. 6 - Purification, Specific Fragmentation, and Separation of Large DNA Molecules. *Recombinant DNA Methodology.* 1989. <https://doi.org/10.1016/B978-0-12-765560-4.50012-5> (ed. R Wu, et al.), pp.139-157. Academic Press, San Diego.
35. Stergachis AB, Debo BM, Haugen E, Churchman LS, Stamatoyannopoulos JA. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science.* 2020;368:1449–1454
36. Turajlic S, Swanton C. Implications of cancer evolution for drug development. *Nat Rev Drug Discov.* 2017;16:441–2.
37. Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, Li B, Arden K, Ren B, Nathanson DA, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature.* 2017;543:122–5.
38. Verhaak RGW, Bafna V, Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer.* 2019;19: 283–288.
39. Wang Y, Wang A, Liu Z, et al. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome research.* 2019;29(8):1329–1342.
40. Wu S, Turner KM, Nguyen N, Raviram R, Erb M, Santini J, Luebeck J, Rajkumar U, Diao Y, Li B, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature.* 2019;575:699–703.
41. Zhu J, Zhang F, Du M, Zhang P, Fu S, Wang L. Molecular characterization of cell-free eccDNAs in human plasma. *Sci Rep.* 2017;7:10968.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

