An Enhancement of Extrachromosomal Circular DNA Enrichment and Amplification to Address the Extreme Low Overlap Between Replicates

Charles M. Burnham[1,2,3=], Alongkorn Kurilung[1,=], Visanu Wanchai[1], Birgritte Regenberg[4], Jesus Delgado-Calle[2], Alexei G. Basnakian[3], Intawat Nookaew[1,2,5,6,*]

[1]Department of Biomedical Informatics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, USA

[2]Department of Physiology and Cell Biology, University of Arkansas for Medical Sciences, Little Rock, AR, USA

[3]Department of Pharmacology and Toxicology, College of Medicine, University of Arkansas for Medical Sciences,

[4] Department of Biology, University of Copenhagen, Copenhagen, Denmark.

[5]Division of Endocrinology, Department Internal Medicine, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, USA

[6]Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

= Shared first author.

*** Correspondence:**

Corresponding Author

INookaew@uams.edu.edu

Keywords: Overlap, Extrachromosomal circular DNA (eccDNA), Exonuclease V, Rolling Circle Amplification (RCA), CRISPR-Cas9, Primer, Long-read sequencing.

Abstract

Extrachromosomal circular DNA (eccDNA) of chromosomal origin is commonly present in all eukaryotic organisms and tissue tested so far. EccDNA populations exhibit immense diversity and a characteristically low degree of overlap between samples, suggesting low inherence of eccDNA between cells or a deficiency the methods by which eccDNA is detected. This study revisits the Circle-seq approach for enrichment of eccDNA to address if these limitations, hypothesizing that experimental procedures significantly contribute to the observed low eccDNA overlap. We optimized the protocol by reducing the time. Linear DNA is digested by increasing exonuclease V activity. We employed CRISPR-Cas9 for mitochondrial linearization, which proved superior to restriction enzymes. A key finding is the critical role of random hexamer primer concentration and genomic DNA input in Rolling Circle Amplification (RCA) for generating high-quality long amplicons from eccDNA (concatemeric tandem copy, CTC), essential for confident *de novo* eccDNA construction from long-read sequencing data. Lower primer concentrations substantially increased the percentage of CTC-derived eccDNA and improved the overlap of identified eccDNAs in technical replicates. Applying this revisited approach to human myeloma and breast cancer cell lines, as well as xenograft models, demonstrated that the optimized conditions enhanced the overlap of detected eccDNA up to over 50% overlap which substantially improved over previous studies (less than 1%). Moreover, Additionally, the oncogenic signature of eccDNAs can be identified across all replicates. These findings provide guidelines for developing standardized procedures for eccDNA profiling, advancing our understanding of eccDNA biology and its potential clinical applications.

Introduction

Extrachromosomal circular DNA (eccDNA) of chromosomal origin represents a fascinating and increasingly studied component of eukaryotic genomes (1-4). These molecules are distinct from the linear chromosomal DNA, existing as separate, covalently closed circles within the cell nucleus. Expression of genes from eccDNA has significant implications for cellular biology, genetics, and disease (1-3). Two defining features of eccDNA populations are their immense diversity and the surprisingly low degree of overlap observed when comparing eccDNA profiles between biological samples and studies (5-12). EccDNAs span a vast size range, from a few hundred base pairs (often termed microDNA) to several megabases (1-3) and form through different DNA repair mechanisms (13) in both soma and germline (7,14). but can carry whole or partial genes, regulatory elements, repetitive sequences, and transposable elements. The specific sequences they contain are highly variable and reflect their genomic origins(1-3).

The eccDNA sequences and the number of copies of a specific eccDNA molecule has been reported to vary dramatically from cell to cell, even within the same tissues (5-7,15-18). This multifaceted diversity strongly indicated that the eccDNA complement within a single cell, or a population of cells, is a complex and dynamic entity. Consequently, the "*low degree of overlap in eccDNA profiles between different samples*" has been reported in many previous studies considering biological replicates of the same tissues (5-9) or cell lines (10-12) and technical replicates based on different approaches to profile eccDNA (9,10). We hypothesize that the current experimental procedure is a key factor contributing to a low degree of overlap due to the high level of heterogeneity of eccDNAs.

Recent advances in enrichment techniques coupled with modern sequencing technologies are aimed to accurately determine the composition of eccDNA. The dominant methodological approach for enrichment is the Circle-Seq method, that involves the treatment of genomic DNA (gDNA) with linear DNA-specific exonucleases to degrade linear chromosomal DNA while enriching circular DNA (19,20). To further purify circular DNA of chromosomal origin, genomic DNA (gDNA) is digested with restriction enzymes (19,20) or CRISPR-Cas9 (21) to linearize and remove the most abundant form of covalently-closed circular DNA in eukaryotic cells,

mitochondrial DNA (mtDNA). Otherwise, these molecules overshadow eccDNA due to their high copy number. Enriched eccDNA is subsequently amplified using rolling circle amplification (RCA) in preparation for high-throughput sequencing. The resulting amplicons are then subject to short-read or long-read sequencing. Computational tools have been developed to resolve eccDNA structures based on the sequence data through the identification of split and discordant read pairs to identify breakpoints where eccDNA circularize, summarized in the recent review (22). Notably, with long-read sequencing, the concatemeric tandem copies (CTCs) derived from RCA contain the total context of an eccDNAs circular conformation and can be captured as single read, yielding the highest confidence in construction of eccDNA molecules (20,23,24).

In this study, we hypothesize that the current experimental procedure to characterize eccDNA is a key factor contributing to the low degree of overlap in eccDNA profiles. we revisited the Circle-seq procedure (19), shortening sample preparation time and significantly increasing the overlap of detected eccDNA across replicates. The time for the eccDNA enrichment procedure was reduced by increasing the frequency of exonuclease V supplementation for linear DNA digestion. Furthermore, the CRISPR-Cas9 approach was utilized to linearize mtDNA. We demonstrated that the concentration of random hexamer primers in RCA plays a crucial role in generating high-quality CTC reads for the *de novo* construction of eccDNA, and the majority of detected eccDNA was derived from these reads leading to an increase in overlapping eccDNA. The level of overlapping was further improved by increasing sampling number. We applied the revisited approach to profiling eccDNA from cultured breast cell lines and human myeloma cells and identified high-overlapping eccDNAs that may play a crucial role in cancer biology.

Results

DNA repair pre-treatment enhances the removal of linearized mtDNA, facilitating the enrichment of circular DNA.

Firstly, we reduced the time of linear chromosomal DNA removal of gDNA from 5 days, based on the original Circle-seq protocol (19), to approximately 6 hours by adding Exonuclease V cocktail every hour (Fig. 1A). Exonuclease V purified from (*Escherichai coli* RecBCD) is used for removal linear DNA is the important step for eccDNA enrichment. The RecBCD is

known to be inhibited by modified DNA and may carry out hydrolysis circular DNA if these contain single strand beaks (25). We then tested whether prior treatment by DNA repair can be used to optimize the removal of linear DNA derived from chromosomes and linearized mitochondria. Using the same DNA pool extracted from HCC1143 cell as a model, we did not observe a difference in the size distribution of unique eccDNAs (Fig 1B) nor number of eccDNA/Mbp (Fig 1C) due to DNA repair pre-treatment. We observed an increased (1.4%, $P < 0.05$) read utilization for eccDNAs construction (Fig 1D), indicating a better digestion of linear DNA in the samples. For mtDNA content after CRISPR-mediated mtDNA removal, we observed an increased removal (0.4%, $P < 0.05$) of the linearized mtDNA when pretreated with the DNA repair enzyme (Fig1E). Based on these results, we incorporated a DNA repair step into our eccDNA enrichment to maximize the removal of mtDNA and enhance eccDNA enrichment. The time required for eccDNA enrichment using our protocol is 15 hours, which is significantly less than the original Circle-Seq procedure which requires 186 hours (See Sup. workflow in details).

Targeted mtDNA linearization by CRISPR-Cas9 outperformed restriction enzyme.

After the chromosomal DNA was removed by ExoV treatment, we aimed to investigate the impact of mtDNA linearization by PmeI restriction enzyme, which is a the most common way, and CRISPR-Cas9 based on sgRNA designed by Feng et al. (21) and our designed sgRNA (see Supp. Figure S1 for the location details of the sgRNAs), followed by removal of linearized mtDNA by ExoV treatment for two hours. We tested removal of mtDNA by qPCR of the mitochondrially encoded Cytochrome C oxidase 1 (*Cox1*), and we observed an over 2-fold of qPCR Ct increase in all conditions compared to internal samples before removal of mtDNA (Supp. Fig. S2A), indicating an efficient removal of mtDNA. Sequencing data revealed that there was no significant difference in percent of mtDNA reads, but there was a noticeable downward trend when considering number of cuts to mtDNA with one, two, and three predicted cuts, respectively Supp. Fig. S2B). The different methods of mtDNA linearization did not result in a difference in neither the number of unique eccDNA identified by long read sequencing and mapping (CReSIL) nor the size distribution of the unique eccDNA (Supp. Fig. S2C and S2D). As expected, there were few PmeI recognition sites harbored on eccDNA after treatment with PmeI, but PmeI eccDNA was retained in the CRISPR-mediated conditions (Fig. 1F top panel).

Interestingly, PmeI eccDNA was larger than non-PmeI eccDNA in the CRISPR-mediated conditions, which is consistent with the probability of the occurrence of a restriction enzyme recognition site increasing the more nucleotides are present (Fig. 1F bottom panel). We further assessed the presence of the recognition site of restriction enzyme PacI, which was used for mtDNA removal in previous studies (20,24). This enzyme has more recognition sites in the human genome compared to PmeI. This experiment resulted in the higher occurrence eccDNA harboring PacI recognition sites, approximately twice more than PmeI in the CRISPR-mediated conditions (Fig. 1G top panel) and, PacI eccDNA was larger than non- PacI eccDNA (Fig. 1G bottom panel).

Increase CTC read generation by RCA Primer titration for high-confidence reconstruction of eccDNA from long read sequences.

RCA products containing CTC provide a high confidence of circular amplified DNA molecules. Using the long-read sequencing approach, we can directly identify the CTC characteristic through CTC read, which yields the highest confidence for reconstructing the eccDNA molecule from the sequencing reads. Based on the eccDNA profiles of the JJN3 cells (Fig. 1B), we observed a strong increase in CTC reads that are used to assign eccDNAs with high confidence (CTC eccDNA) (Fig. 2A). Under standard reaction condition of the RCA by recommended by the manufacture (50 μM), hyperbranching derived from excess random hexamer primer is a typical strategy to obtain a high yield of amplification product. Therefore, we reasoned that if we could increase the incidence of CTC reads by reducing hyperbranching during RCA of eccDNA enriched samples, we could acquire higher-quality reads. We hypothesized that standard reaction conditions from commercially available kits are not optimized for CTC read generation due to the improper ratio of template molecules and substrate random hexametric primers. We then varied the amount of random primer and starting DNA of the same pool of JJN3 cells with two gDNA starting material of 1 μg and 5 μg (recommended by Circle-seq protocol). The amount of RCA product across different amplification condition were correlated with amount of primer concentration as expected (Table S1). Sequencing of the products of the varied conditions revealed that with lower concentrations of primers consistently increasing percent of CTC reads (Fig. 2B) and CTC eccDNA (Fig. 2C). A ~20% increase in CTC eccDNA was observed compared to the standard condition (50 μM) recommended by the kit and

used in many others. The size of CTC eccDNAs is shorter (Fig. 2D) yet their coverage depth is much higher, (Fig. 2E), reflecting higher abundance, as these molecules were preferably amplified. Consequently, the low input for eccDNA enrichment and lower primer concentrations yielded a higher incidence of eccDNA of common origin and high homology (Fig. 3F, Supplemental Figure 6). With the 90% reciprocal overlap cutoff, the ~3.5 % common origin among three technical replicates of 1 μg gDNA with 2.5 μM primer, is substantially improved when compared to previous studies reporting extremely low (less than 1 %) common eccDNA in both technical and biological replicates (5-11). When considering the correlation of eccDNA to gene density by chromosome, we acknowledged higher correlations with lower concentration (Fig. 3G and 3H). Given these observations, we recommend enriching eccDNA from 1 μg of gDNA with 2.5 μM primers for high-abundance eccDNA profiling and enriching eccDNA from 5 μg of gDNA with 5 μM primers for capturing a high diversity of eccDNAs.

Case study 1: EccDNAs profiling of human myeloma OPM-2 cells from culture and xenografted in mice.

We next applied the enhanced protocol to profile the eccDNA of the human myeloma cell line OPM-2 in DNA samples obtained from the digested tibia of xenografted mice. With the limited amount of gDNA extracted from mouse bones (~1 μg of gDNA), which is a mixture of human cancer cells and mouse cells, we applied 1 μg of gDNA with 2.5 μM of primers to profile the eccDNAs. We obtained a wide range of detected human eccDNA, ranging from 17 to 89 eccDNA/Mbp (Fig. 3F), across different 7 samples, which is significantly higher than the values observed in cell culture experiments. We observed 17–30% of CTC reads, which resulted in 63–80% CTC eccDNA (Fig. 3F) across different samples. The distribution of identified eccDNA in different samples was similar (Fig. 3G) and shorter than that in the cell culture experiments (Fig. 3C), possibly due to the amount of starting gDNA. We observed a high correlation of eccDNA to gene density by chromosome. (Supp. Fig S3). We observed that, on average, 50%, 20.7%, 34.4%, 49.5%, 3.7%, and 52.9% of the identified eccDNAs carried CpG, 5' UTR, exon, intron, 3' UTR, and repeats, respectively (Fig. 3H) across the seven samples. The percent overlap of human eccDNAs in xenografted OPM-2 cells (Fig. 3I) was significantly lower compared to cultured OPM-2 cells (Fig. 3E), with an average pairwise overlap of 6.5% and a seven-sample overlap of 0.19%. We arbitrarily selected overlapped eccDNAs found in the three biological

replicates of cultured OPM-2 cells, the seven biological replicates of xenografted OPM-2 cells, and the ten samples combined. One of the selected commonly found eccDNAs in cultured OPM-2 cells was a non-CTC eccDNA (Fig. 3J) with a size of over 15 Kbp and a coverage depth of over 680x, originating from chromosome 11. It carries three exons of *ME3* gene and many repeat elements. Another selected eccDNA is a small CTC eccDNA (~300 bp) (Fig. 3K) with a coverage depth of 14.43x, originating from chromosome 6 was found in common among seven replicates of xenografted OPM-2 samples. We found that this eccDNA contains an exon of the TOMM6 gene with an intron from the *PRICKLE4* gene. Notably, this eccDNA is entirely derived from CpG sequences. Interestingly, we found some common eccDNAs among all the samples, as illustrated in Fig. 3L. One common CTC eccDNA had a size of over 1.2 Kb with a coverage depth of over 50x, originating from chromosome 10 and it carried an exon of the transcription factor *ZEB1* gene.

Case study 2: EccDNAs profiling in Human Breast Cells.

We first performed eccDNA profiles of cultured human breast cancer cells HCC1143, HCC1395, AU565, and SKBR3 using the same procedure as described for profiling cultured OPM-2 cells (5 µg of gDNA with 5 µM primers). We found a very low overlap of less than 5% among biological replicates of all the cancer cells (Fig. 4A), unlike cultured OPM-2 cells (>20% overlap; Fig. 3A). To approach the high heterogeneity of eccDNAs reported in many previous studies, we further increased the number of samples by processing two technical replicates of the same gDNA aliquot then combined them for profiling of eccDNA in the context of biological replicates. We applied 1 µg of gDNA with 2.5 µM of primers to the human breast cancer cell lines, as well as to an epithelial cell line, MCF10A, cultured under similar conditions. Using this strategy, we obtained the average of eccDNA/Mpb with replicates varying from 3.8 to 8.2 eccDNA/Mbp (Fig. 4B). We observed 5–20% of CTC reads (Fig. 4C), which generated 38 – 66% CTC eccDNA (Fig. 4D) across different cell lines. The distribution of the identified eccDNA among the cell lines was slightly different (Fig. 4E). As expected, we observed a high correlation between eccDNA and gene density by chromosome. (Supp. Fig. S4). Assessing the genomic content of eccDNA may provide evidence of its biogenesis. With a similar genomic content across the cells, a significant proportion of eccDNA was enriched for CpG, 5'UTR, exon, and intron, indicating that the biogenesis of eccDNA likely favors regions of high transcriptional

activity (Fig. 4F). As previously reported, eccDNA is enriched by repeat elements, which favors the formation of eccDNA through microhomology (Fig. 4F). In analyzing the breast cell models for eccDNA of both common genomic origin and homology, we first assessed the eccDNA's overlap between two technical replicates over different biological replicates. We observed as high as 18.9% (SKBR3 biological replicate 2) – 50.4% (HCC1143 biological replicate 1) of overlapped eccDNAs (Fig. 4G). The overlap of eccDNAs among the three biological replicates of individual cells was quite high, at 19.5%, 30.1%, 31.8%, 42.2%, and 43.4% for SKBR3, HC1395, AU565, MCF10A, and HC1143, respectively (Fig. 4H). We arbitrarily selected an example of overlapped eccDNA found among the three biological replicates of individual cultured breast cells. We identified a CTC eccDNA (> 1.5 Kbp) (Fig. 4I) with a coverage depth of 70x originating from chromosome 2, which is commonly found among three replicates of non-cancer cell MCF10A. This eccDNA contains an exon of the *BC16143* gene and an intron of the *LRP4* gene, with a long repeat. Another commonly identified non-CTC eccDNA among replicates of AU565 cells had a size of over 9.5 kbp, with a high coverage depth of more than 920x, originating from chromosome 16. We found that this eccDNA contains an exon of the *MYLK3* gene with numerous repeat elements (Fig. 4J). We also found a higher coverage depth (>1900x) of CTC eccDNA as a common feature among replicates of HCC1143 cells (Fig. 4K). This eccDNA had a size of over 1.3 Kbp, which is derived from chromosome 10, and carries an intron-containing microRNA 2 repeats of the *ENTPD1* gene. An overlapped eccDNA among replicates of HCC1143 cells, which had a size over 6Kbp with a coverage depth of 868x derived from chromosome 5 (Fig. 4L), contained 5 repeats, a CpG, and part of the *TCF7* gene. Lastly, we found an over 9.5Kbp non-CTC eccDNA (519x coverage depth), originating from chromosome 15, carrying many exons and introns of the *POLG* gene and many repeats commonly identified among replicates of SKBR3 cells (Fig. 4M).

Oncogenic Signatures of eccDNAs Identified in Multiple Myeloma and Breast Cancer

We further investigated of the curated 758 oncogenes list from COSMIC database (version 102) (26) to see whether their components were carried by the identified eccDNAs derived the samples of the two case studies. As illustrated in Fig. 5A (see supplementary data for details), we observed 8 common oncogenes from the detected 358 oncogenes, distributed in 1557 identified eccDNAs across seven replicates of xenografted OPM-2 cells. For cultured OPM-2 cells, we

observed 52 common oncogenes from the detected 182 oncogenes, distributed in 1557 identified eccDNAs across the three replicates. The breast cells had much smaller number detected oncogenes than the multiple myeloma cells. In the breast cells family, SKBR3 cells have the most detected oncogenes (143 genes) distributed in 326 identified eccDNAs across three replicates and with 21 oncogenes being common among the three replicates. The non-tumorigenic epithelial cell, MCF10A cells had the fewest detected oncogenes of 17, distributed in 60 identified eccDNAs, with 7 of them were in common among the three replicates. Subsequently, the basic characteristic features of eccDNAs carrying oncogenes are illustrated in Fig. 5B (see supplementary data for details). A strong association is clearly seen between the commonly detected oncogenes across replicates, the number of eccDNAs carrying those oncogenes, and the coverage depth in all cells. Interestingly, we found some common oncogenes between xenografted and cultured OPM-2 cells (highlighted in a thick border in Fig. 2B). *NOTCH1,* encoding notch receptor 1, which is a well-known oncogene (27) in multiple myeloma, was detected in all seven replicates of xenografted OPM-2 cells, where it was carried by 40 eccDNAs with a maximum coverage depth of 180x. It was also detected in all three replicates of cultured OPM-2 cells, carried by 7 eccDNAs with an extremely high maximum coverage depth of over 88,000x. *PRDM16*, which encodes a transcription factor that plays a crucial role in various cellular processes, was detected in all seven replicates of xenografted OPM-2 cells, carried by 65 eccDNAs with a very high maximum coverage depth exceeding 2,100x. It was also detected in all three replicates of cultured OPM-2 cells, carried by 17 eccDNAs with a high maximum coverage depth of over 668x. The SKBR3 cells also had PRDM16 as the top-ranked oncogene (bottom right panel). The oncogene was carried by 13 eccDNAs with a high maximum coverage depth of over 864x. Moreover, we also found that *PRDM16* was detected in the rest of the breast cells, appearing in 2, 2, 3, and 2 replicates for AU565, HCC1143, HCC1395, and MCF10A cells, respectively.

Discussion

Our revisited protocol shows that eccDNA enrichment can be conducted in as little as eight hours, a significant reduction compared to traditional protocols involving daily supplementation for up to seven days (19). A rapid enrichment process can potentially benefit clinical applications where timely results are needed. It has been suggested that the removal of mtDNA is necessary

to capture longer eccDNA in colorectal tumor samples (21). Studies indicate that without the removal of mtDNA, upwards of half the reads map to mtDNA (12). Traditional procedures utilize restriction enzymes to linearize mtDNA as part of the eccDNA enrichment process (19,20). This suggests that the removal of mtDNA is an essential step for eccDNA enrichment. Our study determined that longer eccDNAs are more likely to be lost due to treatment with restriction enzymes for the removal of mtDNA. Despite being termed rare-cutting in the context of other restriction enzymes, they still possess a high probability of cutting human chromosomal DNA. Given equal nucleotide content, an 8-mer restriction enzyme is predicted to cut roughly every 65 kilobases. With this probability, it is expected that a longer eccDNA is more likely to contain a recognition site for the restriction enzyme. This is a trend that was observed among eccDNA that includes the recognition site for both PmeI and PacI. Our findings were contrary to a recent study comparing eccDNA enrichment techniques, where the authors suggested eccDNA harboring a recognition site was smaller (12). This was not supported by our findings, where we observed eccDNA containing either a PmeI or PacI recognition site possessing a longer size profile than those that do not contain a recognition site. Given the length of sgRNA used for mtDNA linearization, it is very unlikely that larger eccDNA will contain a target sequence for such a sgRNA. Our results highlight the need to consider the selection of mtDNA linearization techniques. Identifying genes related to genomic stability in breast cancer, such as MTOR, AKT1, and SMAD3, harboring eccDNA, demonstrates the need to screen for genes of interest before selecting a mtDNA linearization technique (See Table S2).

RCA for amplification of eccDNA molecules is a preferred approach, where it creates split and discordant reads used to assemble eccDNA by identifying breakpoints. As a result, protocols and software suites have been developed to exploit this and were recently compared (10). Due to the high processivity of RCA, amplicons of eccDNA can span the circumference of the circle multiple times, resulting in CTC reads. As part of the library preparation for short-read sequencing, these CTC reads are fragmented to meet the read length requirement. In long-read sequencing, the length of the DNA fragment is characteristic of the read length. Therefore, amplicons from RCA are represented in nanopore sequencing data and can be used for the high-confidence identification of eccDNAs. Previous studies did not seek to optimize RCA for the generation of these reads as they could not be detected using short-read sequencing technology.

We found that the manufacturer's recommended concentration of random-hexameric primers is suboptimal for CTC read generation. Given our workflow, a lower random-hexameric primer concentration, paired with lower gDNA input, consistently increased CTC read incidence. We recommend that researchers perform pilot experiments to identify the specific conditions of primer concentrations and input, depending on their availability, as we found in the case studies that different cell lines and conditions have different levels of eccDNAs characteristic and heterogeneity, as previously reported (16).

With a low amount of starting gDNA and low primer in RCA, this can result in over half of eccDNA being derived from CTC reads, as these reads participate in the *de novo* assembly of the majority of eccDNA with high coverage depth. However, the long eccDNAs will be rare compared to the high amount of starting gDNA. Therefore, the amount of gDNA and the amount of primer of choice for RCA will depend on the study's objective, which is to capture a wide range of heterogeneity or high abundance eccDNAs in the considered sample. Nevertheless, our findings revealed a high overlap of identified eccDNA, which can be obtained under proper reaction conditions, unlike previous studies that reported less than 1% overlap (5-12). Within the high degree of eccDNA overlapping achieved by the enhanced procedure, we also found common oncogenes detected across replicates. These common signature oncogenes were carried by many eccDNAs with high coverage depth. This supports the known, strong link between eccDNA and oncogene amplification and overexpression in cancer. (1).

The phenomenon of "low overlap" eccDNA research poses a complex challenge that significantly hinders the consistent detection, characterization, and functional understanding of these diverse genetic elements, as well as the development of biomarkers for clinical diagnostics and therapeutics. This study demonstrated the technical importance of steps in eccDNA enrichment and its application to increase the overlap level of the identified eccDNAs based on long-read sequencing technology. The results from this study provide the basis for the development of a community-accepted procedure for eccDNA profiling in the future, advancing our understanding of eccDNA biology and its implications.

Materials and Methods

Cells and Cell Cultures and Animals.

All human breast cell lines were purchased from ATCC. Human cell lines AU565 (RRID: CVCL_1074), HCC1143 (RRID: CVCL_1245), HCC1395 (RRID: CVCL_1249) were cultured in RPMI-1640 supplemented with 10% FBS. Human cell line SBKR3 (RRID: CVCL_0033) was cultured in McCoy's 5a Modified Medium supplemented with 10% FBS. The human nontumorigenic cell line MCF10A (RRID: CVCL_0598) was propagated and maintained as previously described (28). All cell culture was cultivated in T75 flasks to 70-80% confluency. Cell culture was routinely tested for mycoplasma contaminants using a commercially available PCR kit. (Thermo Fisher Scientific, Waltham, MA, USA). Human OPM-2 multiple myeloma cells (RRID: CVCL_1625) and JJN3 (RRID:CVCL_2078) cells were provided by Dr. Roodman (Indiana University) and Dr. Giuliani (University of Parma, Italy), respectively. The multiple myeloma cells were cultured as previously described (29,30). RPMI 1640 media, fetal bovine serum, Normocin, Plasmocin, and antibiotics (penicillin/streptomycin) were purchased from Invitrogen Life Technologies (Grand Island, NY, USA). For the xenografted animal studies, 7-week-old immunodeficient NSG (RRID:IMSR_JAX:005557) mice were injected intravenously with 10^5 OPM-2 multiple myeloma cells and sacrificed after 3 weeks, as previously reported (31).All procedures involving animals were performed in accordance with the guidelines issued by the University of Arkansas for Medical Sciences Institutional Animal Care and Use Committee (IACUC) (protocol #2022200000489).

Design of guide RNA for Mitochondria Linearization by CRISPR-Cas9.

SgRNAs were designed via CHOPCHOP v3 (32) under the following criteria: 1) Optimal efficiency score (>60), 2) Minimal mismatches, 3) No variants within sgRNA target sequence, 4) all sgRNA share strandedness to avoid creation of exonuclease-resistant fragments. The target-specific spacer sequences for the sgRNA sequences are as follows: sgRNA1: 5'-ACGGTCGGCGAACATCAGTGGGG-3', sgRNA2: 5'- TGTTGAGCCGTAGATGCCGTCGG -3', sgRNA3: 5'- TGAAACCGATATCGCCGATACGG-3'. All sgRNAs were ordered from Integrative DNA Technologies (Coralville, IA, USA).

DNA Purification and EccDNA DNA Enrichment

We performed the Circle-Seq method with the following modifications: DNA purification was conducted using the Monarch Genomic DNA Purification Kit (New England Biolabs, Ipswich, MA, USA) instead of a plasmid miniprep kit. DNA (5 µg or 1 µg when specified) was treated 5 mM NAD^+, 400 units of Tag DNA Ligase (New England Biolabs, Ipswich, MA, USA), and 1X NEB Buffer to a final reaction volume of 50 µL. This reaction was supplemented with 50 units of Exonuclease V (RecBCD) (New England Biolabs, Ipswich, MA, USA), 1 mM ATP, and 1X NEB Buffer 4 to a final reaction volume of 110 µL. Hydrolysis continued at 37°C for 8 hours with the hourly supplementation of 50 units of Exonuclease V, 1 mM of ATP, and 1X NEB Reaction Buffer 4 followed by heat inactivation at 70°C for 30 minutes. Enriched circular DNA was purified by a 1.8x paramagnetic bead (Beckman Coulter, Brea, CA, USA) to volume ratio as previously described (33) Aliquots were retained at this step to serve as internal controls.

Removal of mtDNA via Restriction Endonuclease

For the linearization of mtDNA via restriction enzymes, total enriched circular DNA was treated with 30 units of PmeI (New England Biolabs, Ipswich, MA, USA), 1X rCutSmart Buffer, and nuclease-free water to a final volume of 60 mL. This reaction proceeded for 1 hour at 37°C followed by heat inactivation at 65°C for 20 min. Linearized mtDNA was digested by the addition of two supplements of 20 units of Exonuclease V, 1 mM ATP, and 1X NEB rCutsmart Buffer. Enriched eccDNA was purified by a 1.8x paramagnetic bead (Beckman Coulter, Brea, CA, USA) to volume ratio as previously described. Experiments comparing linearization techniques were taken from the same pooled DNA enrichment.

Removal of mtDNA via CRISPR/Cas9 RNPs

Ribonucleoprotein complexes were formed by the addition of equimolar sgRNA and HiFi Cas9 Nuclease (Integrative DNA Technologies, Coralville, IA, USA). A reaction of 1 µM sgRNA, 1 µM Cas9 Nuclease, 1X NEB rCutsmart Buffer and nuclease-free water were allowed to incubate for 20 min at 25°C to form ribonucleoprotein (RNP) complexes. RNP complexes were then added to total enriched circular DNA in 1X NEB rCutsmart Buffer for a final concentration of 333 nM. This reaction proceeded for 1 hour at 37°C followed heat inactivation for 5 min. at 72°C. Linearized mtDNA was digested by the addition of two supplements of 20 units of Exonuclease V, 1 mM ATP, and 1X NEB rCutsmart Buffer. This reaction proceeded at

37°C for 2 hours, with hourly supplementation, followed by heat inactivation for 30 minutes at 70°C. Enriched eccDNA was purified by a 1.8x paramagnetic bead (Beckman Coulter, Brea, CA, USA) to volume ratio.

Rolling Circle Amplification of eccDNA

Enriched eccDNA amplified by Phi29 polymerase amplification using the phi29-XT RCA Kit according to the manufacturer's protocol. (New England Biolabs, Ipswich, MA, USA). Random hexameric primers were diluted for RCA optimization, and incubation was extended to 4 hours. eccDNA amplicons (5µg) were treated with 50 units of T7 endonuclease in 1X NEB Buffer 2 to resolve cruciform structures. This reaction was incubated for 2 hours at 37°C followed by 1.8x paramagnetic bead cleaning.

Acquisition, Base Calling, Data Preprocessing, and Identification of eccDNA Long Read Data

EccDNA amplicons (2.5 µg) from each sample were input for sequencing library preparation using the Native Barcoding Kit (Oxford Nanopore Technologies, Oxford, United Kingdom) according to the manufacturer's protocol. Equivalent amounts of barcoded DNA were pooled to ~1500 ng and used as input for the Ligation Sequencing Kit (ONT) according to the manufacturer's protocol. Experiments using DNA purified from breast cancer cell lines were sequenced on R9.4 FLO-MIN106D flow cells on a MinION (Mk1B; ONT). Experiments involving CTC read generation, human breast cancer cell profiling, and MM cell lines and xenografts were sequenced on R10.4.1 FLO-MIN106D flow cells using a MinION (Mk1B; ONT). Fast5/POD5 files were demultiplexed and basecalled using Guppy software version 6 or Dorado software version via the high accuracy model (dna_r9.4.1_450bps_hac_cfg) and (dna_r10.4.1_e8.2_400bps_hac) to generate fastq files. Native barcode adaptors were trimmed using porechop and reads <200 bp were filtered by NanoFilt (34) The resulting fastq files were processed with CReSIL[16] for eccDNA identification under the following parameters: default settings for the trimming step, relaxed settings for the identification step (-minrsize 50 -break -1), and finally default settings for the identification step.

Restriction Endonuclease Recognition Site Mapping on eccDNA

Consensus fasta sequences of eccDNA were extracted from CReSIL output. Sequences were loaded using the SeqIO module of the Biopython Library. Sequences were concatenated to confer the circularity and restriction enzyme recognition sites were searched using the Restriction module of Biopython, which can be modified for other restriction endonucleases. Results were compiled in a .txt format, containing the ID of the eccDNA, as well as the site and position of the site. This output was cross-referenced to CReSIL output based on eccDNA ID, where a positive hit based on id populated a new column for the status of the restriction endonuclease. This method was tested on mtDNA, where it correctly identified the recognition site of PmeI and PacI.

Identification of CTC-derived eccDNA

Reads that contained concatemers of eccDNA were identified in the output of the CReSIL trim step. Unique CTC read IDs were extracted and used to filter fastq files using seqtk subseq. Resulting filtered fastq files was processed with CReSIL where BED coordinates of CTC-derived eccDNA were compared by a within sample basis using bedtools intersect querying for IDs that possessed a 90% reciprocal overlap (bedtools intersect -r -f 0.9). These IDs were cross referenced to original CReSIL output based on ID.

Overlapping Analysis of identified eccDNAs

All eccDNA genomic coordinates were concatenated into a single master BED file that were subject to comparison, serving as a reference set for all unique eccDNA to be compared. To quantify overlap between each replicate and the master set, we ran bedtools annotate -both optionz, with -i as the master set and -files as individual replicates. Intervals were retained for a given replicate only if all coverage fractions were ≥ 0.90, enforcing a 90 % overlap, where these values were used to create a boolean presence/absence matrix indexed by unique interval ID.

Data availability: All of sequencing data of 79 samples were deposited and publicly available at NCBI SRA database under BioProject PRJNA1283158. The update of CRSIL software to compatible with R10 chemistry of Oxford Nanopore sequencing is available at https://github.com/visanuwan/cresil.

Funding.

Musculoskeletal (MSK) Health and Disease Creativity Hub of the UAMS College of Medicine, P20GM125503, T32GM106999, R37CA251763

References

1. Noer JB, Horsdal OK, Xiang X, Luo Y, Regenberg B. Extrachromosomal circular DNA in cancer: history, current knowledge, and methods. Trends Genet **2022**;38:766-81
2. Zhao Y, Yu L, Zhang S, Su X, Zhou X. Extrachromosomal circular DNA: Current status and future prospects. Elife **2022**;11
3. Li R, Wang Y, Li J, Zhou X. Extrachromosomal circular DNA (eccDNA): an emerging star in cancer. Biomark Res **2022**;10:53
4. Paulsen T, Kumar P, Koseoglu MM, Dutta A. Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells. Trends in Genetics **2018**;34:270-8
5. Henriksen RA, Jenjaroenpun P, Sjøstrøm IB, Jensen KR, Prada-Luengo I, Wongsurawat T*, et al.* Circular DNA in the human germline and its association with recombination. Molecular Cell **2022**;82:209-17.e7
6. Liang X, Arrey G, Qin Y, Alvarez-Gonzalez L, Hariprakash JM, Ma J*, et al.* EccDNA atlas in male mice reveals features protecting genes against transcription-induced eccDNA formation. Nat Commun **2025**;16:1872
7. Moller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P*, et al.* Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. Nat Commun **2018**;9:1069
8. Sin STK, Jiang P, Deng J, Ji L, Cheng SH, Dutta A*, et al.* Identification and characterization of extrachromosomal circular DNA in maternal plasma. Proc Natl Acad Sci U S A **2020**;117:1658-65
9. Zole E, Hansen LB, Haskó J, Gerovska D, Araúzo-Bravo MJ, Noer JB*, et al.* Technical and biological variations in the purification of extrachromosomal circular DNA (eccDNA) and the finding of more eccDNA in the plasma of lung adenocarcinoma patients compared with healthy donors. bioRxiv **2024**:2024.03.05.583481
10. Gao X, Liu K, Luo S, Tang M, Liu N, Jiang C*, et al.* Comparative analysis of methodologies for detecting extrachromosomal circular DNA. Nat Commun **2024**;15:9208
11. dos Santos CR, Hansen LB, Rojas-Triana M, Johansen AZ, Perez-Moreno M, Regenberg B. Variation of extrachromosomal circular DNA in cancer cell lines. Computational and Structural Biotechnology Journal **2023**;21:4207-14
12. Lu W, Li F, Ouyang Y, Jiang Y, Zhang W, Bai Y. A comprehensive analysis of library preparation methods shows high heterogeneity of extrachromosomal circular DNA but distinct chromosomal amount levels reflecting different cell states. Analyst **2023**;149:148-60

13. Eugen-Olsen RAB, Hariprakash JM, Oestergaard VH, Regenberg B. Molecular mechanisms of extrachromosomal circular DNA formation. Nucleic Acids Res **2025**;53
14. Henriksen RA, Jenjaroenpun P, Sjostrom IB, Jensen KR, Prada-Luengo I, Wongsurawat T, *et al.* Circular DNA in the human germline and its association with recombination. Mol Cell **2022**;82:209-17 e7
15. Denoth-Lippuner A, Krzyzanowski MK, Stober C, Barral Y. Role of SAGA in the asymmetric segregation of DNA circles during yeast ageing. Elife **2014**;3
16. Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, *et al.* Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature **2017**;543:122-5
17. Koo DH, Molin WT, Saski CA, Jiang J, Putta K, Jugulam M, *et al.* Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed Amaranthus palmeri. Proc Natl Acad Sci U S A **2018**;115:3332-7
18. Sinclair DA, Guarente L. Extrachromosomal rDNA circles--a cause of aging in yeast. Cell **1997**;91:1033-42
19. Møller HD. Circle-Seq: Isolation and Sequencing of Chromosome-Derived Circular DNA Elements in Cells. Methods in molecular biology (Clifton, NJ) **2020**;2119:165-81
20. Wang Y, Wang M, Zhang Y. Purification, full-length sequencing and genomic origin mapping of eccDNA. Nature Protocols 2022 18:3 **2022**;18:683-99
21. Feng W, Arrey G, Zole E, Lv W, Liang X, Han P, *et al.* Targeted removal of mitochondrial DNA from mouse and human extrachromosomal circular DNA with CRISPR-Cas9. Comput Struct Biotechnol J **2022**;20:3059-67
22. Li F, Ming W, Lu W, Wang Y, Dong X, Bai Y. Bioinformatics advances in eccDNA identification and analysis. Oncogene **2024**;43:3021-36
23. Wanchai V, Jenjaroenpun P, Leangapichart T, Arrey G, Burnham CM, Tümmler MC, *et al.* CReSIL: accurate identification of extrachromosomal circular DNA from long-read sequences. Briefings in Bioinformatics **2022**;23:1-11
24. Wang Y, Wang M, Djekidel MN, Chen H, Liu D, Alt FW, *et al.* eccDNAs are apoptotic products with high innate immunostimulatory activity. Nature 2021 599:7884 **2021**;599:308-14
25. Leung WY, Chung LH, Kava HW, Murray V. RecBCD (Exonuclease V) is inhibited by DNA adducts produced by cisplatin and ultraviolet light. Biochem Biophys Res Commun **2018**;495:666-71
26. Sondka Z, Dhir NB, Carvalho-Silva D, Jupe S, Madhumita n, McLaren K, *et al.* COSMIC: a curated database of somatic variants and clinical data for cancer. Nucleic Acids Research **2024**;52:D1210-D7
27. Sabol HM, Ferrari AJ, Adhikari M, Amorim T, McAndrews K, Anderson J, *et al.* Targeting notch inhibitors to the myeloma bone marrow niche decreases tumor growth and bone destruction without gut toxicity. Cancer research **2021**;81:5102-14
28. Basnakian AG, Apostolov EO, Yin X, Abiri SO, Stewart AG, Singh AB, *et al.* Endonuclease G promotes cell death of non-invasive human breast cancer cells. Exp Cell Res **2006**;312:4139-49

29. Sabol HM, Ashby C, Adhikari M, Anloague A, Kaur J, Khan S, *et al.* A NOTCH3-CXCL12-driven myeloma-tumor niche signaling axis promotes chemoresistance in multiple myeloma. Haematologica **2024**;109:2606-18
30. Sabol HM, Ferrari AJ, Adhikari M, Amorim T, McAndrews K, Anderson J, *et al.* Targeting Notch Inhibitors to the Myeloma Bone Marrow Niche Decreases Tumor Growth and Bone Destruction without Gut Toxicity. Cancer Res **2021**;81:5102-14
31. Delgado-Calle J, Anderson J, Cregor MD, Condon KW, Kuhstoss SA, Plotkin LI, *et al.* Genetic deletion of Sost or pharmacological inhibition of sclerostin prevent multiple myeloma-induced bone disease without affecting tumor growth. Leukemia **2017**;31:2686-94
32. Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res **2019**;47:W171-W4
33. Feng W, Arrey G, Zole E, Lv W, Liang X, Han P, *et al.* Targeted removal of mitochondrial DNA from mouse and human extrachromosomal circular DNA with CRISPR-Cas9.
34. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics **2018**;34:2666-9

Figure Legends

Fig 1. Impact of DNA repair pre-treatment on eccDNA enrichement using HCC1143 cells as a model. A.) Bar plot showing mean raw Ct of COX5 with varying gDNA input and time of digestion (*n* = 3). B). Violin plot of size distribution of samples with and without DNA repair. C.) Bar plot showing mean eccDNA/Mb with a and without DNA repair (*n* = 3) D.) Bar plot showing mean percent read utilization for eccDNA construction with and without DNA repair (*n* = 3). E.) Bar plot showing mean percent reads mapping to mtDNA with and without DNA repair. F.) Bar plot showing mean percent of eccDNA harboring PmeI cutsite and associated violin plots of eccDNA with (Orange) and without (Blue) PmeI cutsite (n=3). G.) Mean percent of eccDNA harboring PacI cutsite and associated violin plots of eccDNA with (Orange) and without (Blue) PmeI cutsite (n=3). (Student's T-test where statistics indicated.)

Fig 2. Influence of RCA primer titration to eccDNA generation using JJN3 cells as a model. A.) Box and whisker plot showing distribution of and correlation between CTC eccDNA percentage and CTC read percentage (n = 3). B.) Bar plot showing CTC read percentage at varying primer concentrations using 1 µg input (Pink) and 5 µg input (Blue) (*n* = 3). B.) Bar plot showing percent CTC eccDNA at varying primer concentrations under 1 µg input (Pink) and 5 µg input

(Blue) (*n* = 3). D.) Violin plot showing length of CTC eccDNA (Green) and non-CTC eccDNA (Purple) using 1 µg input or 5 µg input. D.) Violin plot showing coverage depth of CTC eccDNA (Green) and non-CTC eccDNA (Purple) using 1 µg input or 5 µg input. F.) Bar plot showing percent overlap of replicates at varying primer concentrations using 1 µg input (Pink) and 5 µg input (Blue) (*n* = 3). H.) Correlation between eccDNA per Mb and Coding Gene per Mb with varied primer concentration using 5 µg (shades of Blue) and 1 µg (shades of Pink), points represent mean between triplicates. Statistical analysis was performed with Pearson's correlation. RCA: Rolling Circle Amplification.

Fig 3. Comparison of eccDNA characteristics between OPM-2 myeloma cell cultured and xenografted. A.) Graphic depicting 7-week-old immunodeficient NSG mice injected intravenously with 10^5 OPM-2 MM cells. B.) Bar graphs showing mean eccDNA/Mb, mean CTC read percentage, and mean CTC eccDNA of in vitro cultured OPM-2 cells (*n* = 3). C.) Violin plot showing length of eccDNA *in vitro* cultured OPM-2 cells. D.) Half moon chart showing percent eccDNA carrying features of in vitro cultured OPM-2 cells (*n* = 3). E.) Bar plot showing percent overlap of eccNDA between 2 and 3 replicates of in vitro cultured OPM-2 cells. F.) Bar graphs showing mean eccDNA/Mb, mean CTC read percentage, and mean CTC eccDNA of OPM-2 xenograft (*n* = 7) G.) Violin plot showing length of eccDNA of OPM-2 xenografted. H.) Half moon chart showing percent eccDNA carrying features of OPM-2 xenografted (*n* = 7). I.) Bar plot showing percent overlap of eccDNA between of all permutations of OPM-2 xenografted (*n* = 7). Bar plot showing percent overlap of eccDNA between of all permutations of OPM-2 xenografted (*n* = 7). Examples of overlapped eccDNAs identified from cultured OPM-2 cells (J), xenografted OPM-2 cells (K) and all of them (L). See lanes annotation in Sup. Fig S5.

Fig 4. eccDNA profiles across different human breast cancer cell lines. A.) Bar graph showing percent overlap between replicates of breast cancer cell lines. B.) Bar plot showing mean eccDNA/Mb of breast cancer cell lines (*n* = 3). C.) Bar plot showing mean CTC read percentage of breast cancer cell lines (*n* = 3). D.) Bar plot showing mean CTC eccDNA percentage of breast cancer cell lines (*n* = 3) E.) Violin plot showing average length of eccDNA of breast cancer cell lines (*n* = 3). F.) Bar plot showing mean percentage of eccDNA carrying selected features in breast cancer cell lines (*n* = 3). G.) Bar plot showing mean overlap between combinations of

technical replicates of breast cancer cell lines. H.) Bar plot showing overlap between 3 biological replicates of breast cancer cell lines. Examples of overlapped eccDNAs identified from MCF10A cells (I), AU565 cells (J), HCC1143 cells (K), HCC1395 cells (L) and SKBR3 cells (M). See lanes annotation in Sup. Fig S5.

Fig 5 Oncogene distribution analysis of identified eccDNAs from the two case studies (Figs. 3 and 4). A) Bar plots show the number of oncogenes, or their components detected on eccDNAs identified across a number of replicates of different cells. B) Bubble plots show the basic characteristics (size on the y-axis and coverage depth on the x-axis) of eccDNAs carrying an oncogene or its components. The dot size represents the number of eccDNAs that carry a specific oncogene across different replicates. The top-ranked oncogenes (large dots) found in all replicates are indicated by name tags with a gray arrow. The figure annotations are the same for all individual plots. In all plots, the color indicates the number of replicates in which an oncogene was detected on the identified eccDNAs.

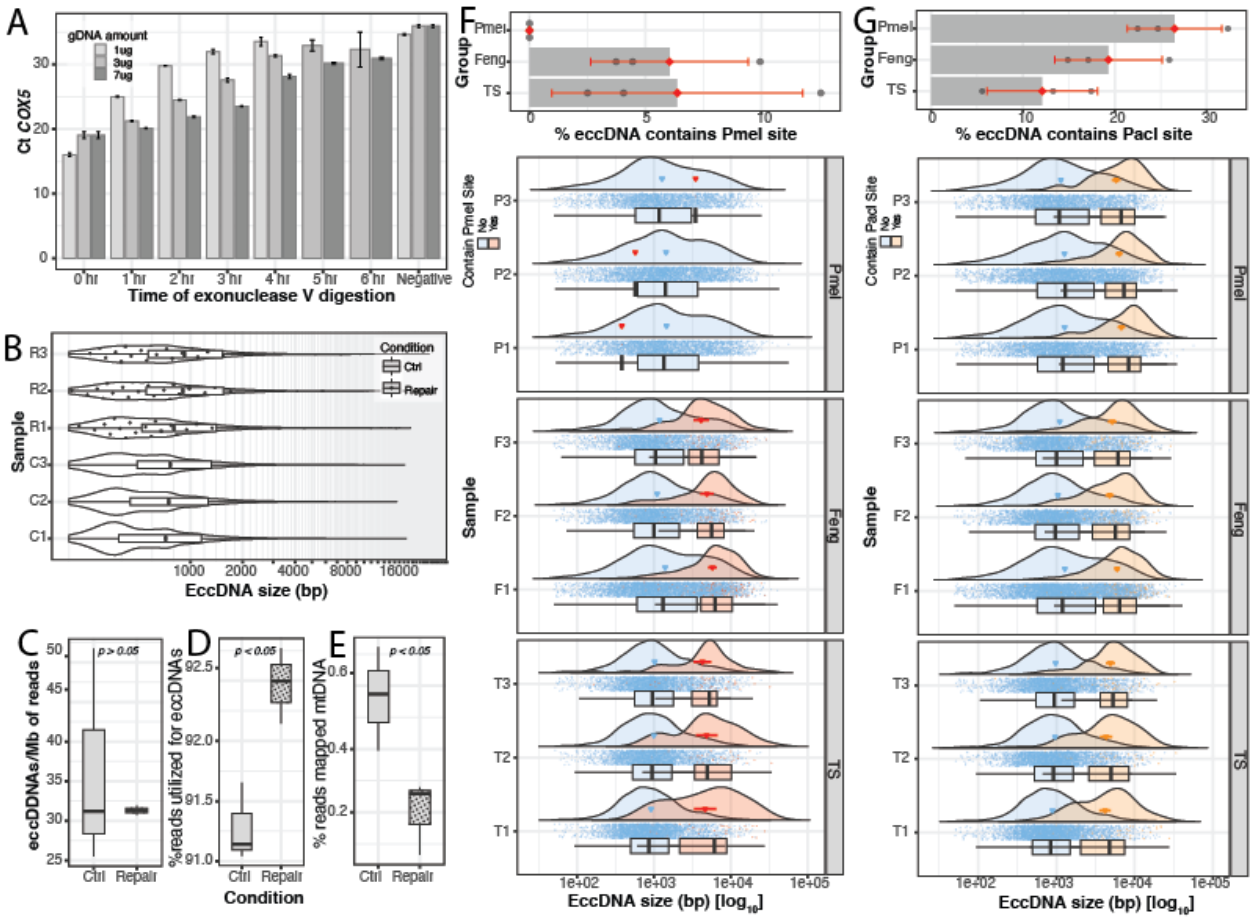
A
gDNA amount
1ug
3ug
7ug
Ct COX5
Time of exonuclease V digestion
0 hr
1 hr
2 hr
3 hr
4 hr
5 hr
6 hr
Negative
B
Condition
Ctrl
Repair
Sample
EccDNA size (bp)
C
eccDDNAs/Mb of reads
p > 0.05
D
%reads utilized for eccDNAs
p < 0.05
E
% reads mapped mtDNA
p < 0.05
Condition
F
Group
Pmel
Feng
TS
% eccDNA contains Pmel site
Contain Pmel Site
No
Yes
EccDNA size (bp) [log10]
G
% eccDNA contains PacI site
Contain PacI Site

Fig1

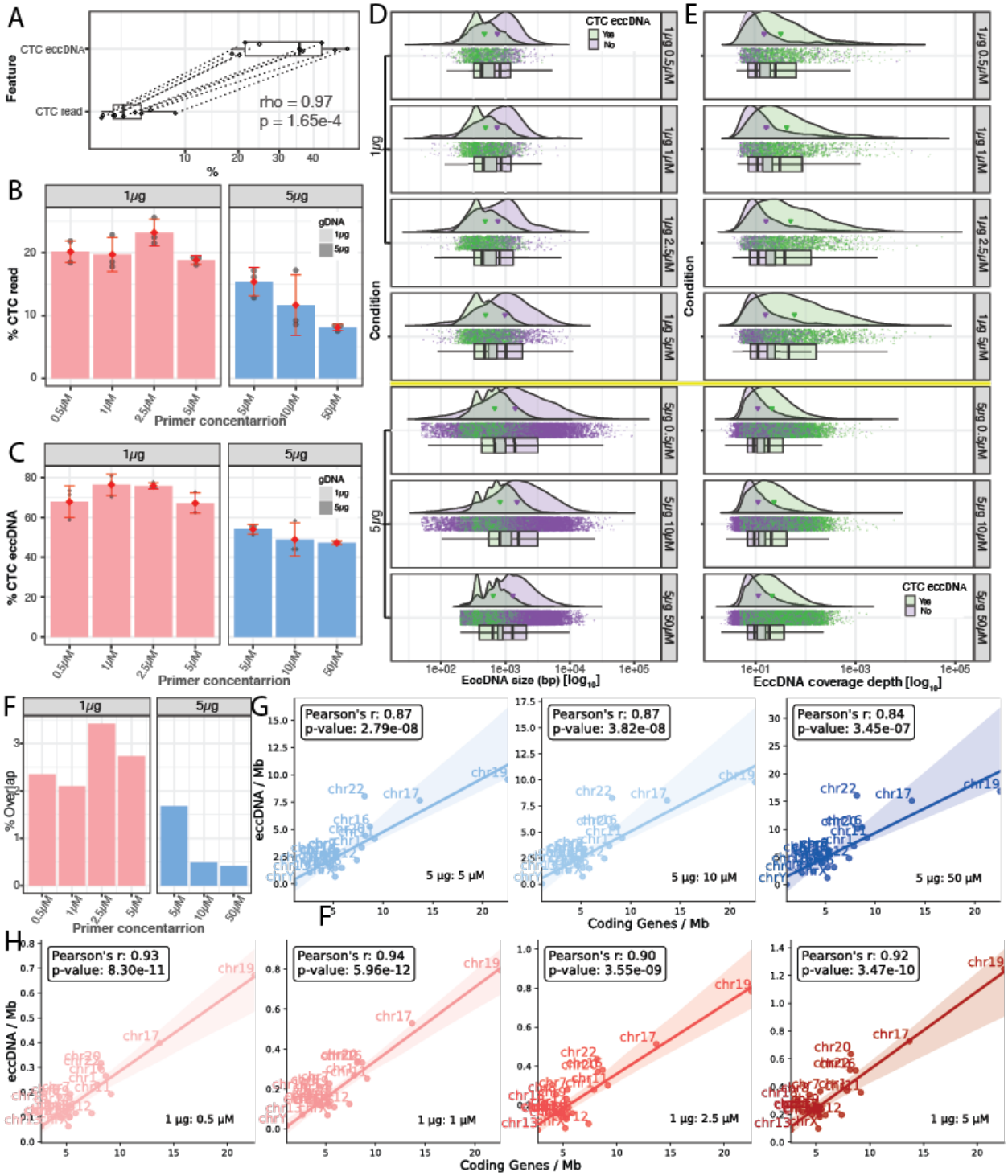
A
Feature
CTC eccDNA
CTC read
rho = 0.97
p = 1.65e-4
%
B
1μg
5μg
gDNA
% CTC read
Primer concentarrion
C
% CTC eccDNA
D
CTC eccDNA
Yes
No
Condition
EccDNA size (bp) [log10]
E
EccDNA coverage depth [log10]
F
% Overlap
G
Pearson's r: 0.87
p-value: 2.79e-08
5 μg: 5 μM
Pearson's r: 0.87
p-value: 3.82e-08
5 μg: 10 μM
Pearson's r: 0.84
p-value: 3.45e-07
5 μg: 50 μM
eccDNA / Mb
Coding Genes / Mb
H
Pearson's r: 0.93
p-value: 8.30e-11
1 μg: 0.5 μM
Pearson's r: 0.94
p-value: 5.96e-12
1 μg: 1 μM
Pearson's r: 0.90
p-value: 3.55e-09
1 μg: 2.5 μM
Pearson's r: 0.92
p-value: 3.47e-10
1 μg: 5 μM

Fig 2

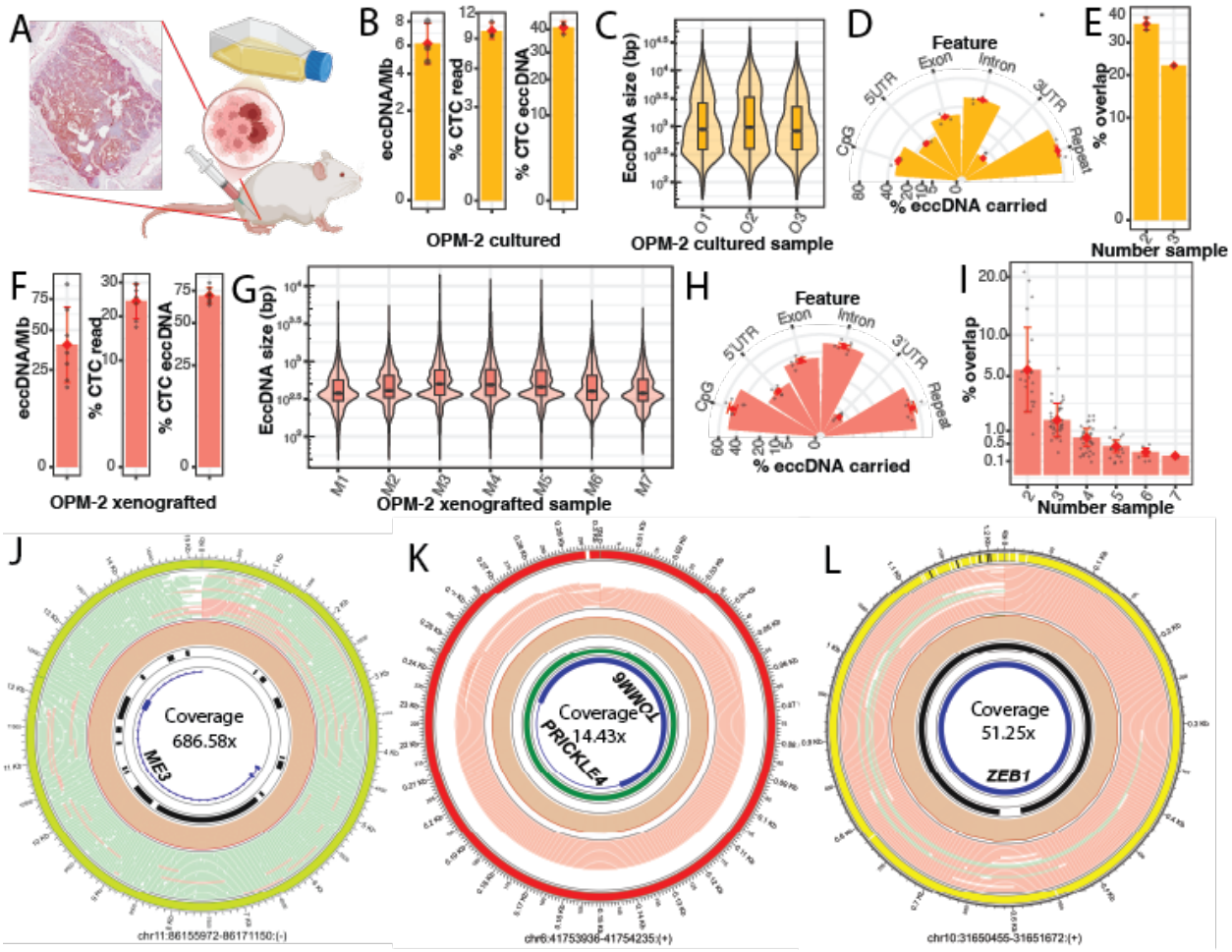
A
B
eccDNA/Mb
% CTC read
% CTC eccDNA
OPM-2 cultured
C
EccDNA size (bp)
O1
O2
O3
OPM-2 cultured sample
D
Feature
Exon
Intron
5'UTR
3'UTR
CpG
Repeat
% eccDNA carried
E
% overlap
Number sample
F
eccDNA/Mb
% CTC read
% CTC eccDNA
OPM-2 xenografted
G
EccDNA size (bp)
M1
M2
M3
M4
M5
M6
M7
OPM-2 xenografted sample
H
Feature
Exon
Intron
5'UTR
3'UTR
CpG
Repeat
% eccDNA carried
I
% overlap
Number sample
J
Coverage
686.58x
ME3
chr11:86155972-86171150:(-)
K
Coverage
14.43x
PRICKLE4
TOMM6
chr6:41753936-41754235:(+)
L
Coverage
51.25x
ZEB1
chr10:31650455-31651672:(+)

Fig 3

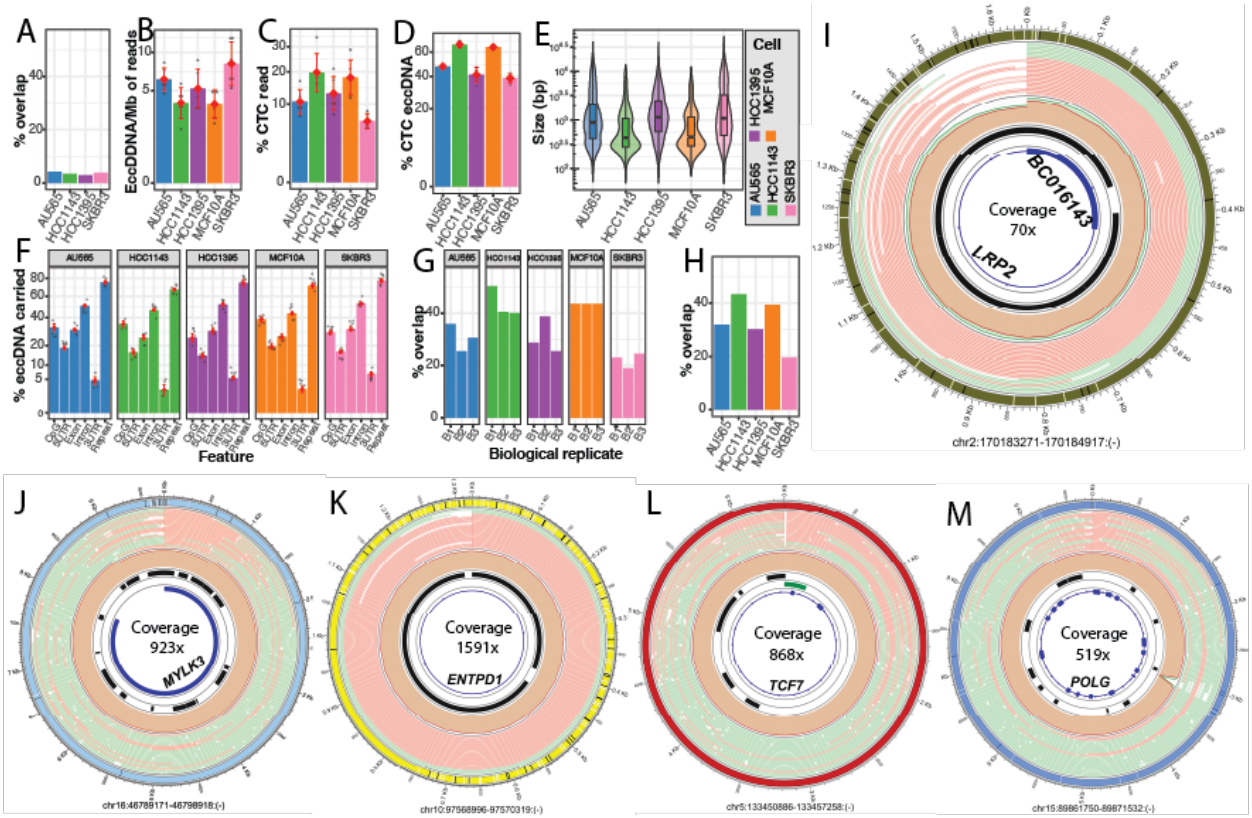
A
% overlap
B
EccDNA/Mb of reads
C
% CTC read
D
% CTC eccDNA
E
Size (bp)
Cell
HCC1395
MCF10A
AU565
HCC1143
SKBR3
F
% eccDNA carried
Feature
G
% overlap
Biological replicate
H
% overlap
I
BC016143
Coverage
70x
LRP2
chr2:170183271-170184917:(-)
J
Coverage
923x
MYLK3
K
Coverage
1591x
ENTPD1
L
Coverage
868x
TCF7
M
Coverage
519x
POLG

Fig 4

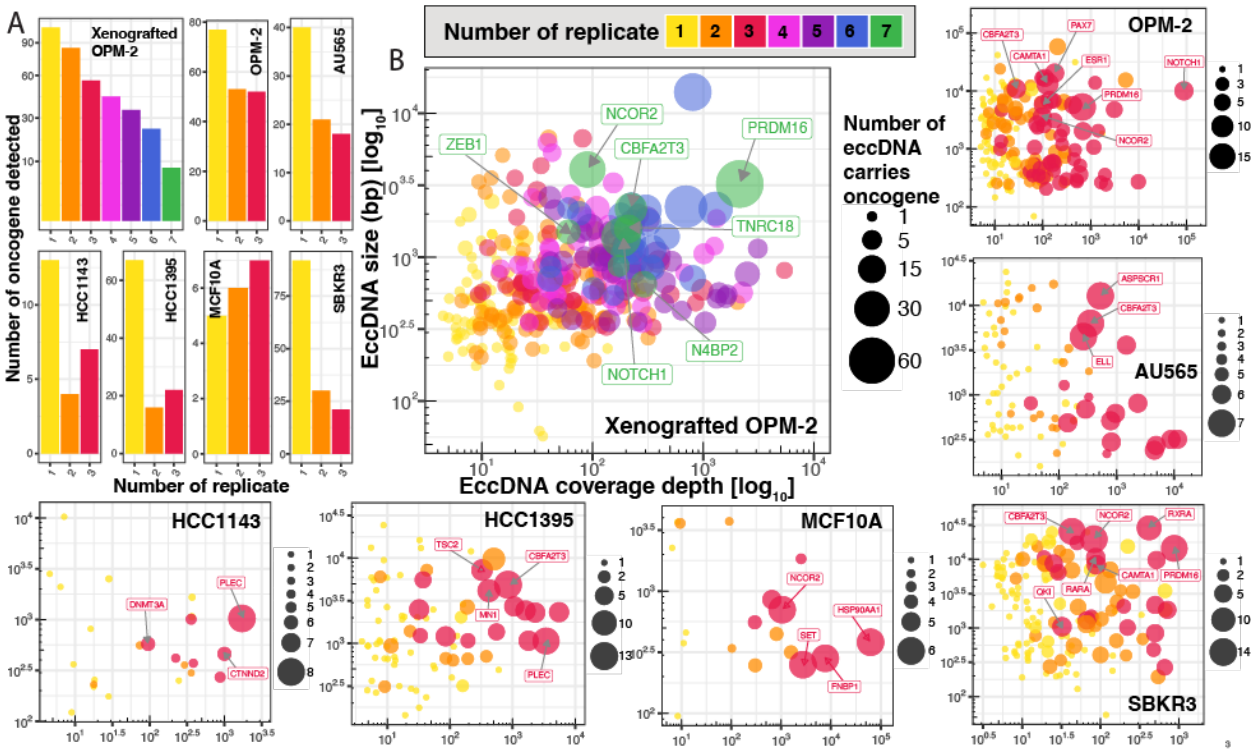
A
Xenografted OPM-2
OPM-2
AU565
HCC1143
HCC1395
MCF10A
SBKR3
Number of oncogene detected
Number of replicate
B
Number of replicate
1
2
3
4
5
6
7
EccDNA size (bp) [log10]
EccDNA coverage depth [log10]
Xenografted OPM-2
ZEB1
NCOR2
CBFA2T3
PRDM16
TNRC18
N4BP2
NOTCH1
Number of eccDNA carries oncogene
1
5
15
30
60
OPM-2
CBFA2T3
PAX7
CAMTA1
ESR1
NOTCH1
PRDM16
NCOR2
AU565
ASPSCR1
CBFA2T3
ELL
HCC1143
PLEC
DNMT3A
CTNND2
HCC1395
TSC2
CBFA2T3
MN1
PLEC
MCF10A
NCOR2
HSP90AA1
SET
FNBP1
SBKR3
CBFA2T3
NCOR2
RXRA
QKI
RARA
CAMTA1
PRDM16

Fig 5