Electronic Supplementary Material (ESI) for Analyst.

A comprehensive analysis of library preparation methods shows high heterogeneity of extrachromosomal circular DNA but distinct chromosomal amount levels reflecting different cell states

1. Supplementary tables and figures

Supplementary Table 1. Number of identified eccDNA by different methods.

Sample ID	Total mapped reads	mtDNA content (%)	eccDNA number (Junctional tag >1)	Identical junction site	eccDNAs with 90% overlap of sequences	eccDNA number (Junctional tag ⩾1)
S-M-R-1	44509774	4.16	4614	28	174	32887
S-M-R-2	69804373	3.83	6412			55234
S-P-R-1	65869445	0.13	5091	106	213	51250
S-P-R-2	57227911	0.20	5278			32486
S-N-R-1	66139198	69.94	2003	194	210	7642
S-N-R-2	100442167	78.24	2704			9821
S-non-R-1	63909412	66.70	749	0	5	7790
S-non-R-2	52860791	61.51	1989			11265
Tn5-M-R-1	117325408	2.43	9996	55	230	52470
Tn5-M-R-2	110636048	1.94	8816			43530
Tn5-P-R-1	101089780	0.046	14871	45	172	53507
Tn5-P-R-2	68096033	0.074	13421			45754
Tn5-N-R-1	124984899	48.37	6966	29	97	20989
Tn5-N-R-2	97204441	59.85	7004			20688
Tn5-non-R-1	113601795	52.10	5312	18	83	24606
Tn5-non-R-2	114435314	46.39	6666			30066
Tn5-M-1	39962730	6.75	386	10	12	31575
Tn5-M-2	42562810	6.17	204			19888
Tn5-P-1	44833118	2.33	178	5	6	17077
Tn5-P-2	43915152	1.53	276			15137
Tn5-N-1	35764335	87.70	1	0	0	1085
Tn5-N-2	37139503	86.76	4			837
Tn5-non-1	36794114	85.29	1	0	0	195
Tn5-non-2	36025804	83.40	0			114

Supplementary Table 2. Number of genes harbor by eccDNA.

Sample ID	Number of genes harbor by eccDNA (Junctional tag >1)	Overlap genes number
S-M-R-1	295	32
S-M-R-2	455	
S-P-R-1	533	2
S-P-R-2	383	
S-N-R-1	100	1
S-N-R-2	53	
S-non-R-1	103	0
S-non-R-2	119	
Tn5-M-R-1	550	1
Tn5-M-R-2	536	
Tn5-P-R-1	979	9
Tn5-P-R-2	471	
Tn5-N-R-1	479	2
Tn5-N-R-2	591	
Tn5-non-R-1	358	0
Tn5-non-R-2	313	
Tn5-M-1	17	0
Tn5-M-2	4	
Tn5-P-1	7	7
Tn5-P-2	125	
Tn5-N-1	0	0
Tn5-N-2	0	
Tn5-non-1	0	0
Tn5-non-2	0	

Supplementary Table 3. Number of identified eccDNA by Flec.

Sample ID	eccDNA number	eccDNA/Mb reads	mtDNA content (%)	Multiple fragments eccDNA number
ONT-Flec-P-1	16988	1724.24	0.02	2851
ONT- Flec-P-2	19221	1523.97	0.04	3207
ONT- Flec-non-1	6233	776.72	56.83	319
ONT- Flec-non-2	8361	900.46	53.93	456

Supplementary Table 4. Number of full passes reads.

Sample ID	Total mapped reads	≥1 full passes reads	Percentage of ≥1 full passes reads	≥2 full passes reads	Percentage of ≥2 full passes reads
ONT-Flec-P-1	9852438	251228	2.55%	128263	1.30%
ONT-Flec-P-2	12612434	318534	2.53%	159776	1.27%
ONT-Flec-non-1	8024724	75423	0.94%	23747	0.30%
ONT-Flec-non-2	9285234	108846	1.17%	39462	0.42%

Supplementary Table 5. Number of identified eccDNA by CReSIL.

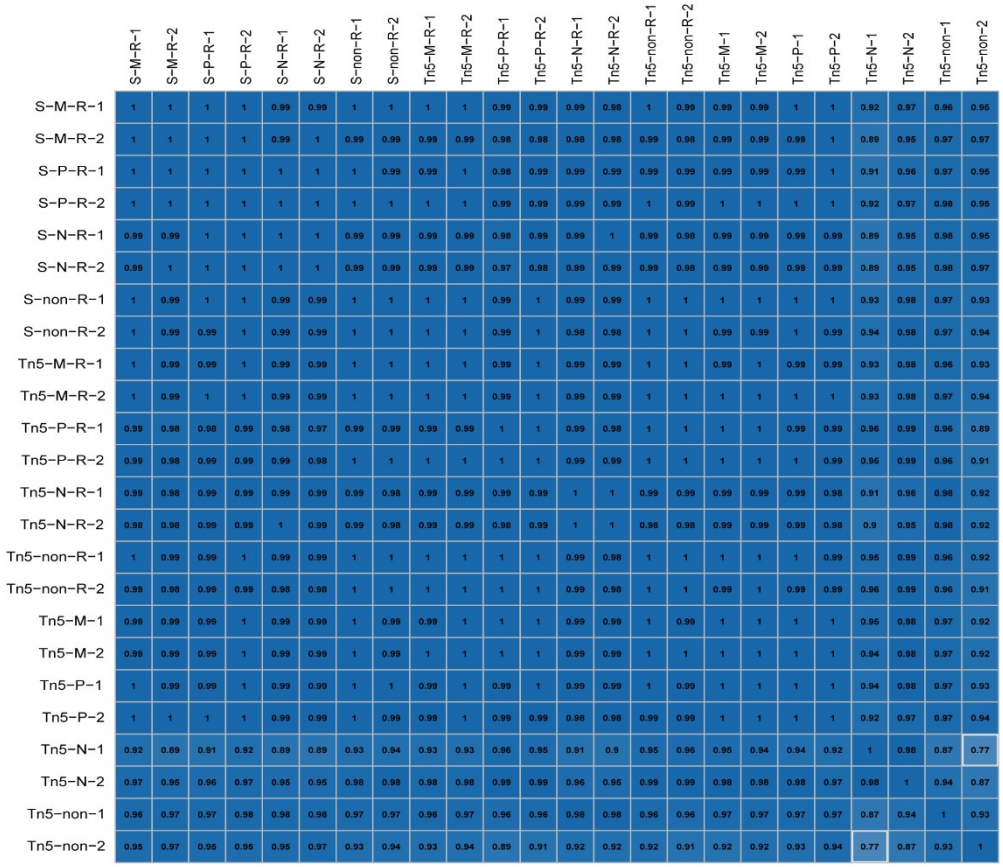
Sample ID	eccDNA number	Multiple fragments eccDNA number
CReSIL-P-1	22649	7810
CReSIL-P-2	25889	9246
CReSIL-non-1	14377	8650
CReSIL-non-2	15576	8740

Supplementary Table 6. Comparison of different bioinformatics tools.

	ONT-P-1		ONT-P-2		ONT-non-1		ONT-non-2	
Method	single fragment eccDNA number	eccDNAs with 90% overlap of sequences	single fragment eccDNA number	eccDNAs with 90% overlap of sequences	single fragment eccDNA number	eccDNAs with 90% overlap of sequences	single fragment eccDNA number	eccDNAs with 90% overlap of sequences
Flec	14137	8192	16014	8949	5914	1788	7905	2484
CReSIL	14839		16643		5727		6836	

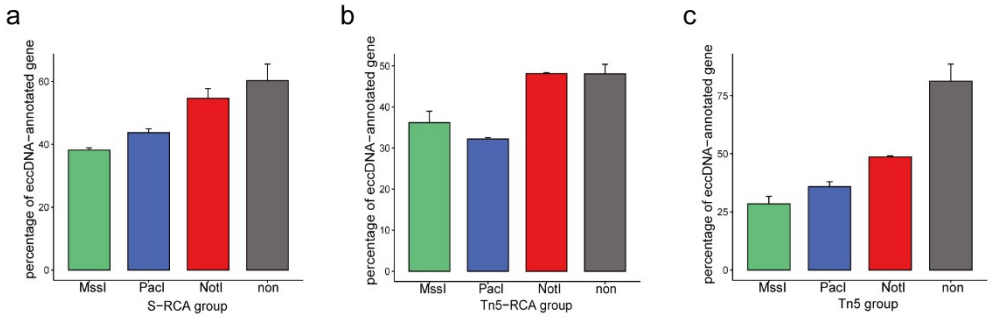
Supplementary Table 7. Number of identified eccDNA in different cell lines.

Cell lines	Total mapped reads	eccDNA number	number of eccDNAs with 90% overlap of sequences in GES-1
GES-1	72691769	7828	7828
HepG2	51760640	9083	44
HL7702	68848781	3714	11
MDA-MB-453	84996581	32969	33
MCF-12A	77703215	1021	5



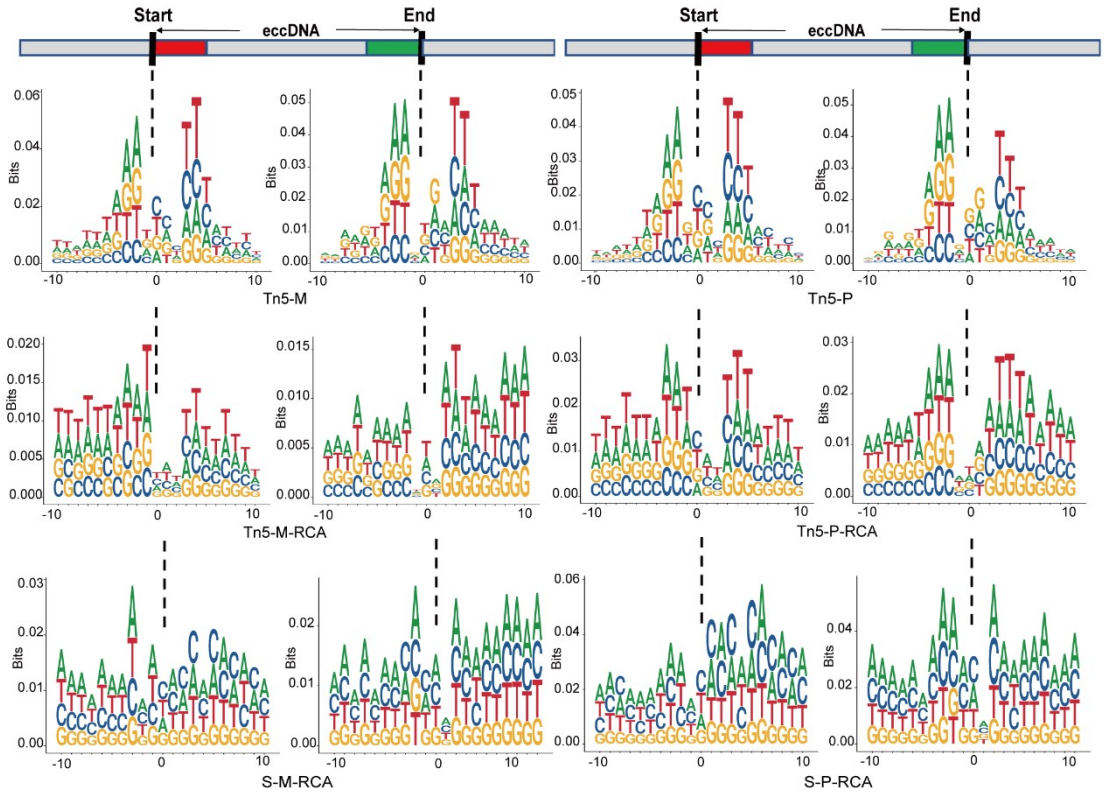
Supplementary Figure 1. Pearson correlation analysis of the distribution of eccDNA in different genomic elements in all treatment samples.

S, sonication; Tn5, Tn5 transposase tagmentation; M, MssI restriction enzymes; P, PacI restriction enzymes; N, NotI restriction enzymes; non, non-digested; R, rolling-circle amplification (RCA).

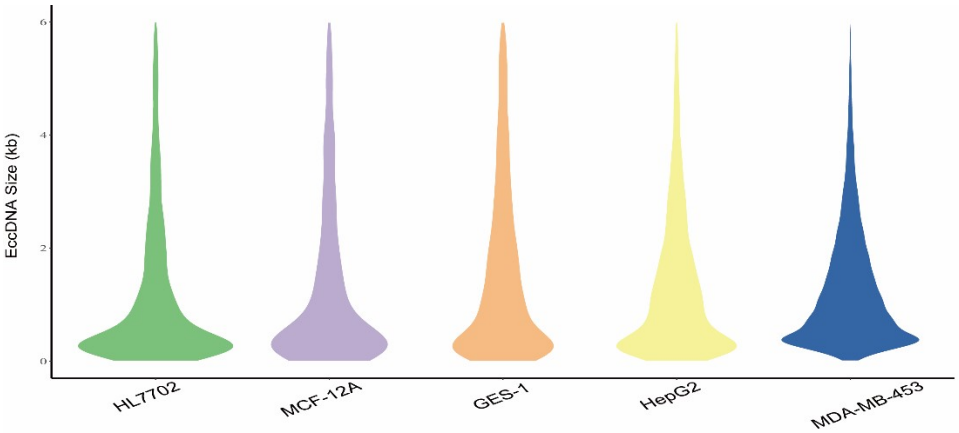


Supplementary Figure 2. A comparison of the percentage of eccDNA-annotated genes in different groups.

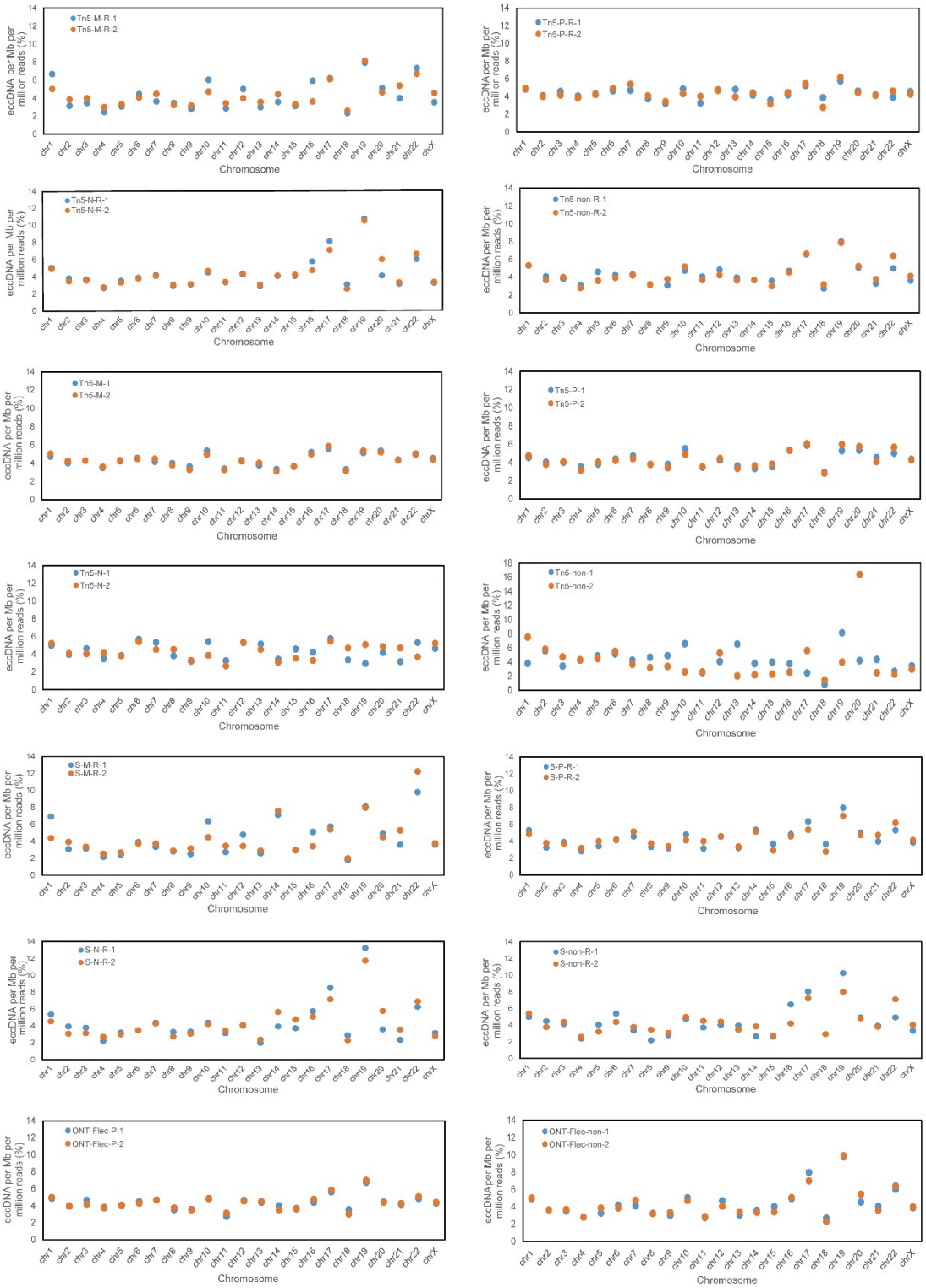
S, sonication; Tn5, Tn5 transposase tagmentation.



Supplementary Figure 3. Comparison of junctional nucleotide motif patterns in each replicate samples. The typical pattern of a pair of high frequency trinucleotid segments with 4-bp "spacers" in non-RCA replicate samples (Tn5-M and Tn5-P), while RCA samples were atypical.
S, sonication; Tn5, Tn5 transposase tagmentation; M, MssI restriction enzymes; P, PacI restriction enzymes.

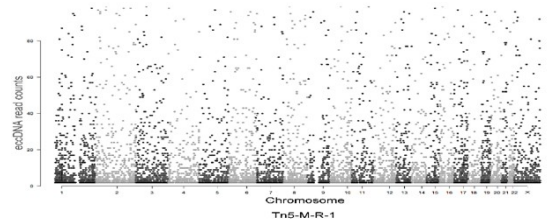


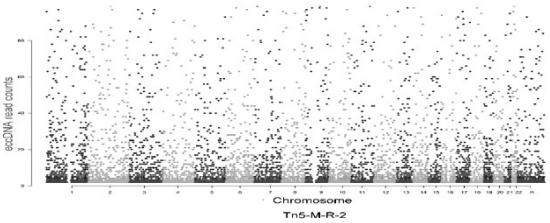
Supplementary Figure 4. Violin plots depicting the length distribution of eccDNA in different cell lines.

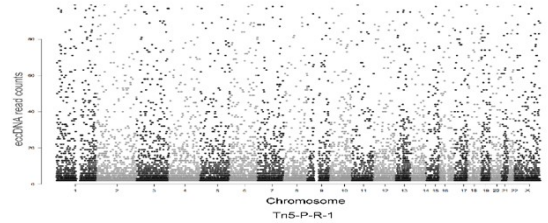


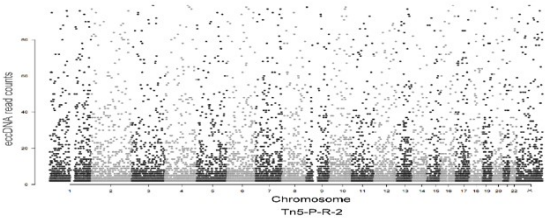
Supplementary Figure 5. Comparison of normalized counts of eccDNAs on each chromosome in all treatment samples.

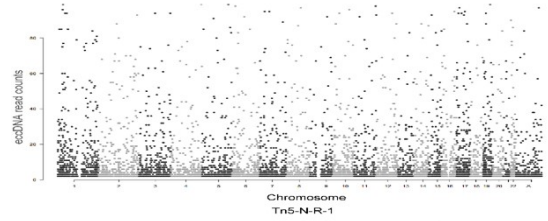
S, sonication; Tn5, Tn5 transposase tagmentation; M, MssI restriction enzymes; P, PacI restriction enzymes; N, NotI restriction enzymes; non, non-digested; R, rolling-circle amplification (RCA); ONT, Oxford Nanopore Technologies.

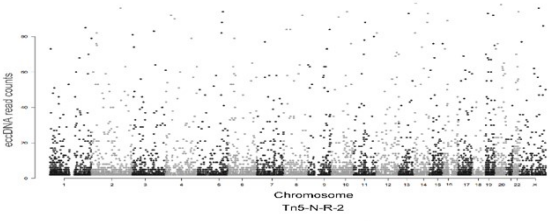

eccDNA read counts
Chromosome
Tn5-M-R-1

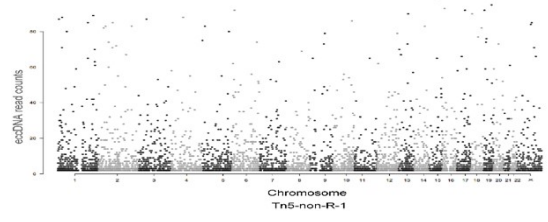

eccDNA read counts
Chromosome
Tn5-M-R-2

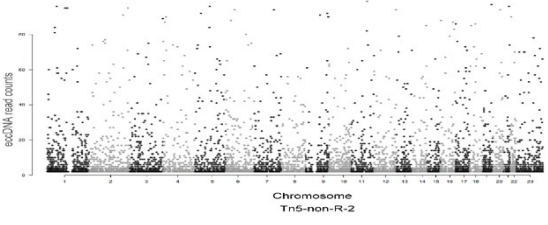

eccDNA read counts
Chromosome
Tn5-P-R-1

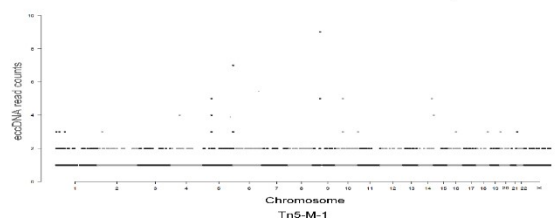

eccDNA read counts
Chromosome
Tn5-P-R-2

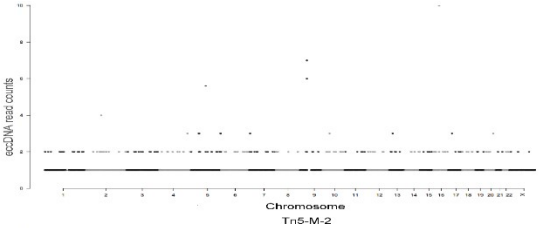

eccDNA read counts
Chromosome
Tn5-N-R-1

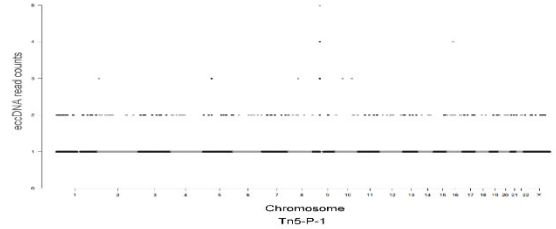

eccDNA read counts
Chromosome
Tn5-N-R-2

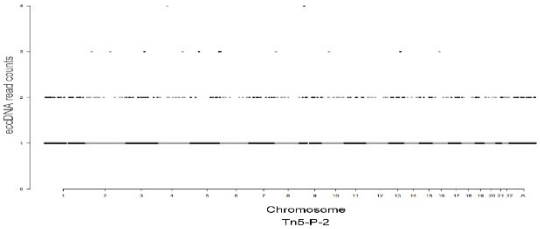

eccDNA read counts
Chromosome
Tn5-non-R-1

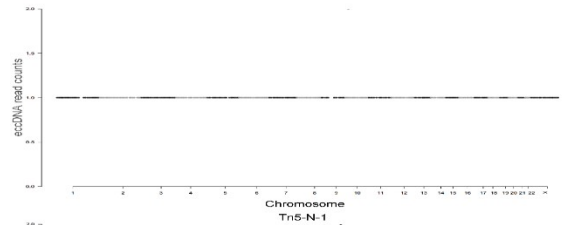

eccDNA read counts
Chromosome
Tn5-non-R-2

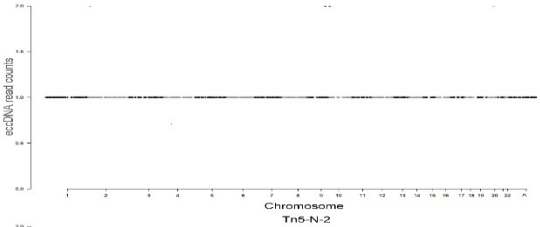

eccDNA read counts
Chromosome
Tn5-M-1

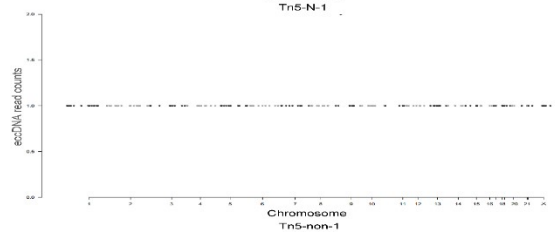

eccDNA read counts
Chromosome
Tn5-M-2

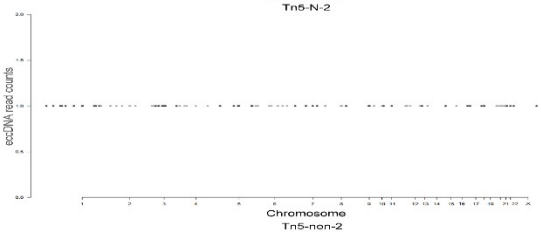

eccDNA read counts
Chromosome
Tn5-P-1

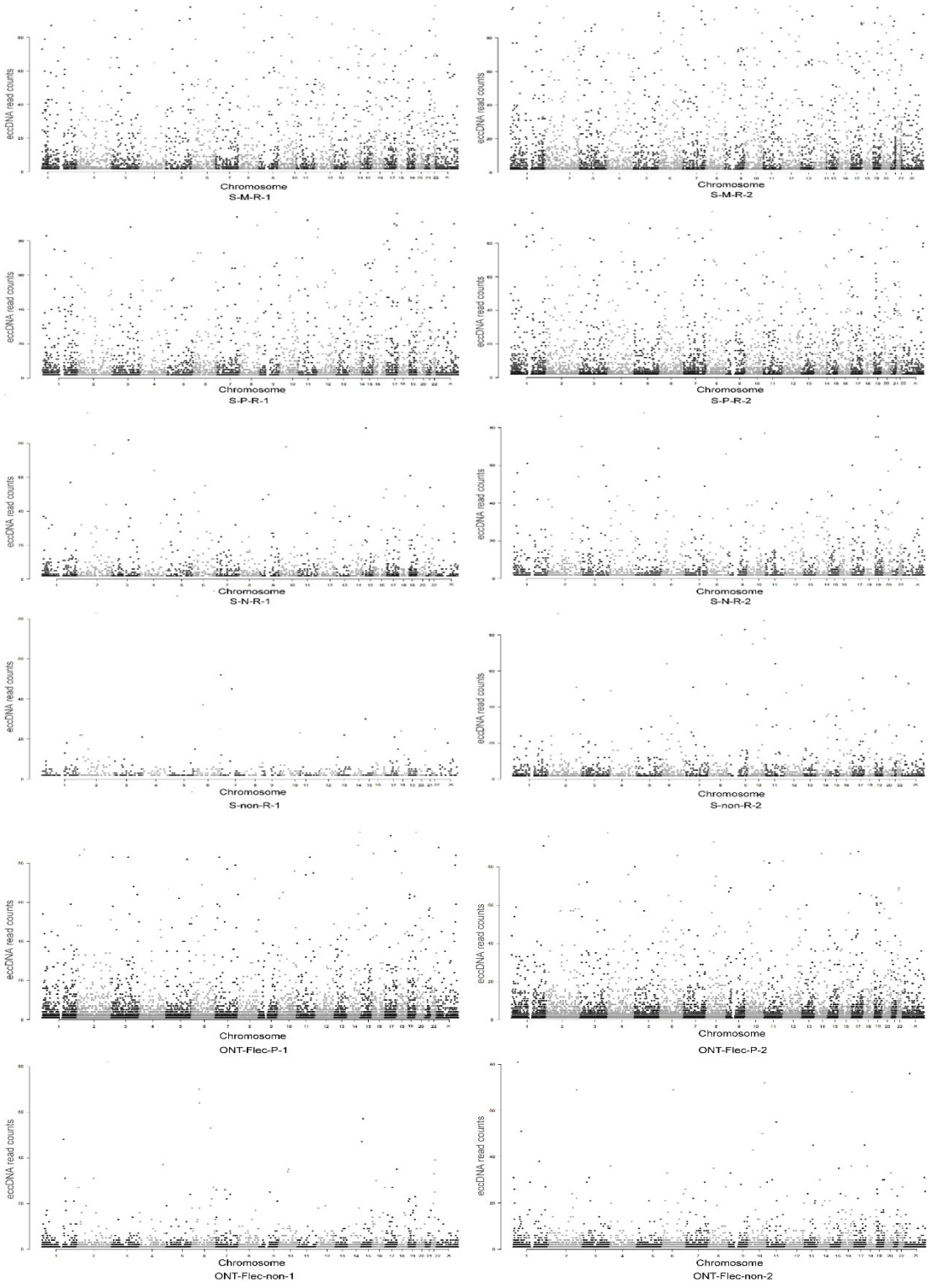

eccDNA read counts
Chromosome
Tn5-P-2


eccDNA read counts
Chromosome
Tn5-N-1


eccDNA read counts
Chromosome
Tn5-N-2


eccDNA read counts
Chromosome
Tn5-non-1


eccDNA read counts
Chromosome
Tn5-non-2



Supplementary Figure 6. Manhattan plot of the distribution of eccDNAs across chromosomes in all treatment samples.

S, sonication; Tn5, Tn5 transposase tagmentation; M, MssI restriction enzymes; P, PacI restriction enzymes; N, NotI restriction enzymes; non, non-digested; R, rolling-circle amplification (RCA); ONT, Oxford Nanopore Technologies.

2. Supplementary Methods

2.1 EccDNA detecting method for cell types experimental

Genomic DNA was isolated from different cell types using the Allprep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions. To purify eccDNAs, Plasmid-Safe ATP-dependent DNase (Epicentre) was used to digest linear DNA at 37 °C for 5 days, with the addition of ATP and DNase every 24 hours. To confirm the elimination of linear chromosomal DNA, we performed a quantitative polymerase chain reaction (qPCR) to amplify a chromosomal marker to evaluate linear chromosomal DNA following exonuclease digestion.

The purified eccDNAs were used as templates for rolling-circle amplification. A 20 μl reaction was set up using 1 mM dNTPs, 10 U Phi29 DNA polymerase (New England Biolabs), 50 μM Exo-resistant random primer (Thermo Fisher), 0.02 U inorganic pyrophosphatase (Thermo Fisher), and 1× Phi29 DNA polymerase buffer. The reaction was performed at 30 °C for 18 h and purified using AMPure XP beads (Beckman). The amplified DNA was sheared by sonication (Covaris) and then the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) was used to prepare the library. The Illumina Novaseq 6000 platform was used to sequence these libraries in 150 bp paired-end mode.

The sequencing reads were assessed by FastQC and filtered by Trimmomatic. Then the clean data were aligned to the human reference genome (GRCh38) using the default settings of BWA-MEM. The Circle-map software was then used to identify eccDNA with parameters of "Circle-Map Realign -i alignment/sample.bam -qbam alignment/sample_qname.bam -sbam alignment/sample_coord.bam -fasta refFa".

2.2 Bioinformatics tools and parameters

For NGS data was aligned by BWA-MEM with parameters of "bwa mem <indxbase> <in1.fq> <in2.fq>".

eccDNA identified by Circle_finder tools and downloaded from https://github.com/pk7zuva/Circle_finder/blob/master/circle_finder-pipeline-bwa-mem-samblaster.sh with parameters of "bash circle_finder-pipeline-bwa-mem-

samblaster.sh 10 hg38.fa in1.fastq in2.fastq 10 sampleID hg38".

For cell lines sequencing data, Circle-Map tools (https://github.com/iprada/Circle-Map) was used to identified eccDNA with parameters of "Circle-Map Realign -i alignment/sample.bam -qbam alignment/sample_qname.bam -sbam alignment/sample_coord.bam -fasta refFa".

Nanopore sequencing data was aligned by minimap2 with parameters of "-ax map-ont –c --secondary=no". eccDNA calling used Flec tools and downloaded from https://github.com/YiZhang-lab/eccDNA_RCA_nanopore with parameters of "./eccDNA_RCA_nanopore.py --fastq mapping/sample_name.fastq.gz --paf mapping/sample_name.paf --info <info.tsv> --seq <seq.fa> --var <var.tsv> --reference <path/to/reference.fa> --verbose | tee <out.log>".

As comparison, CReSIL tools (https://github.com/visanuwan/cresil) was also used to identify eccDNA with parameters of "cresil trim -t 8 -fq sample.fq -r reference.mmi" and "cresil identify -t 8 -minrsize 40 -depth 1 -break 1 -fa reference.fa -fai reference.fa.fai -fq sample.fq -trim sample/trim.txt".

BedTools was used to genomic element annotation with parameters of "intersectBed -a sample.bed -b element.bed -wa". The parameters of "bedtools getfasta –fi refFa –bed sample.bed –fo sample.fa" was used to extract sequences. The parameters of "bedtools intersect -f 0.90 -r -wa -wb -a sample1.bed -b sample2.bed" was used to identify eccDNA with 90% overlap of sequences.