Supplementary Information

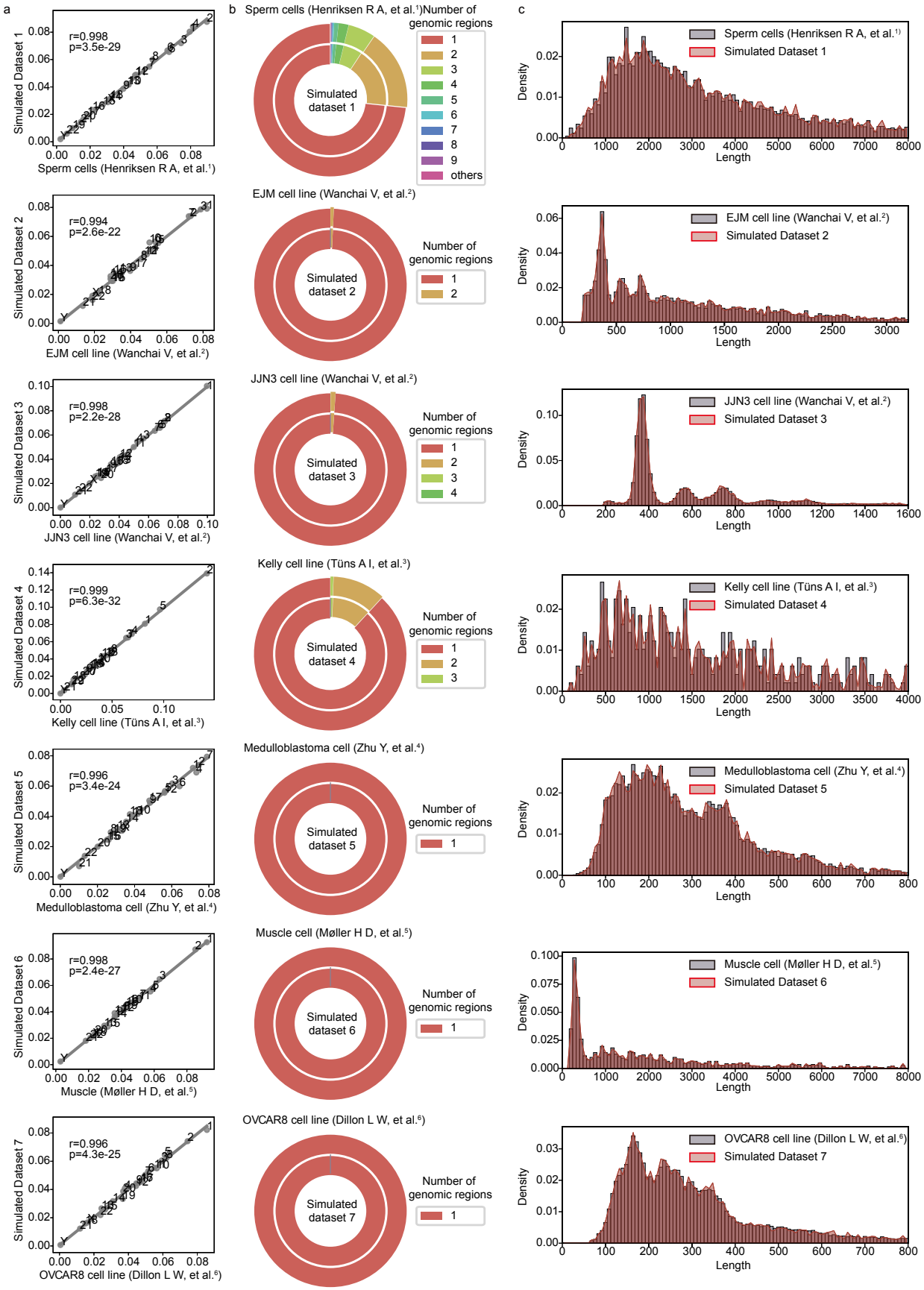
Comparative analysis of methodologies for detecting extrachromosomal circular DNA

Xuyuan Gao[1,7], Ke Liu[1,7], Songwen Luo[1,7], Meifang Tang[1,2], Nianping Liu[1], Chen Jiang[1,2], Jingwen Fang[1,3], Shouzhen Li[1], Yanbing Hou[1], Chuang Guo[1,4,5*], Kun Qu[1,2,6*]

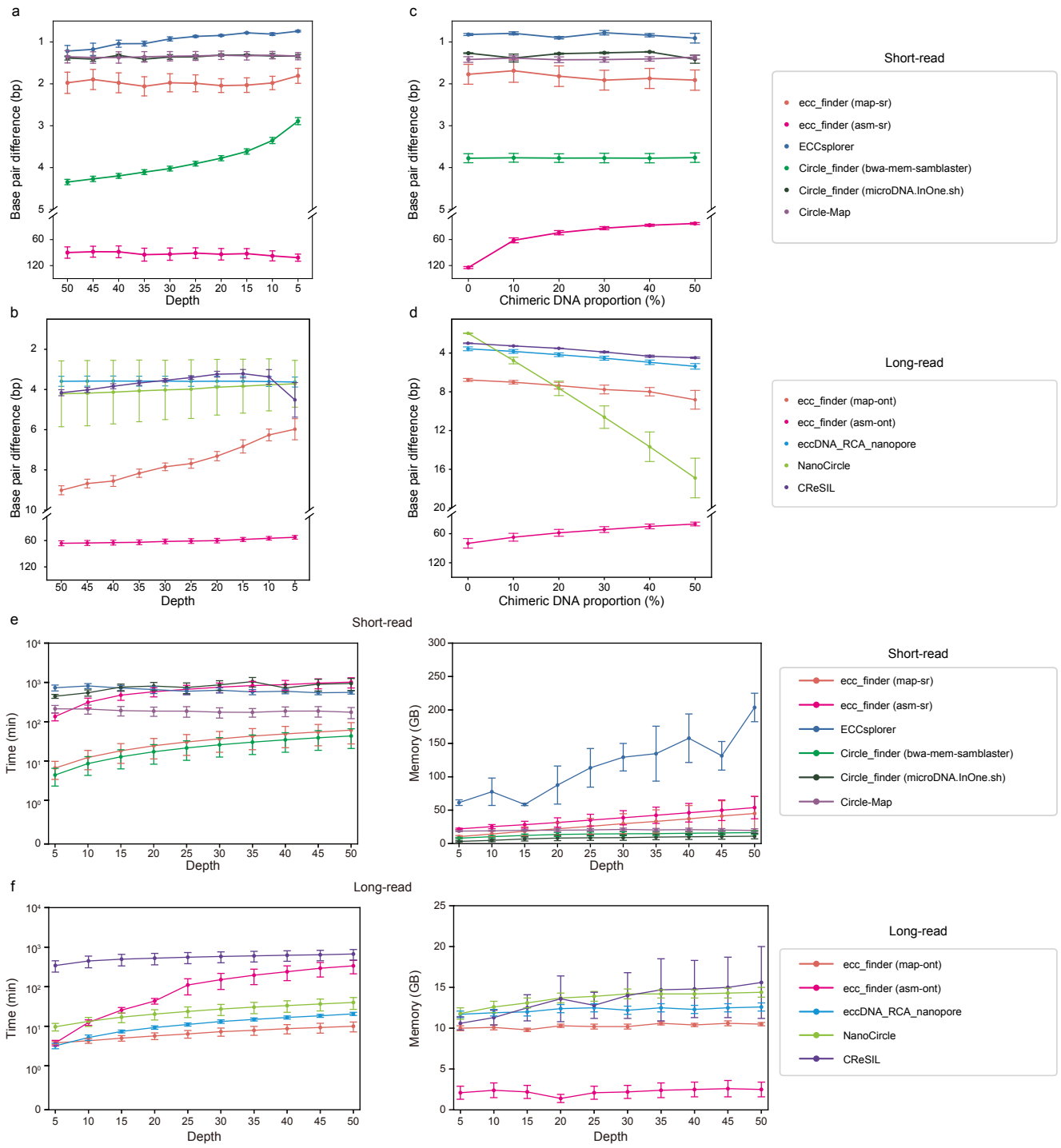
[1]Department of Oncology, The First Affiliated Hospital of USTC, School of Basic Medical Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China
[2]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China.
[3]HanGene Biotech, Xiaoshan Innovation Polis, Hangzhou, Zhejiang, China, 31200
[4]School of Pharmacy, Bengbu Medical University, Bengbu, China, 233030.
[5]Department of Rheumatology and Immunology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China, 230021
[6]School of Biomedical Engineering, Suzhou Institute for advanced Research, University of Science and Technology of China, Suzhou, China, 215000
[7]These authors contributed equally
[*]Corresponding should be addressed to Kun Qu (qukun@ustc.edu.cn); Chuang Guo (gchuang@ustc.edu.cn)

Dataset ID	Sample	Tissue/ Disease	Sequencing Platform	Pipeline	Number of Simple eccDNA	Number of Chimeric eccDNA
Dataset 1	Sperm cells	Semen	Nanopore	NanoCircle	8306	3014
Dataset 2	EJM cell line	Multiple myeloma	Nanopore	CReSIL	4635	33
Dataset 3	JJN3 cell line	Multiple myeloma	Nanopore	CReSIL	90661	988
Dataset 4	Kelly cell line	Neuroblastoma	Nanopore	eccDNA_RCA_nanopore	431	58
Dataset 5	Medulloblastoma	Medulloblastoma	Illumina	Circle-Map	12518	/
Dataset 6	Muscle cells	Muscle	Illumina	ecc_finder (map-sr)	1421	/
Dataset 7	OVCAR8 cell line	Ovarian cancer	Illumina	Circle_finder (bwa-mem-samblaster)	57231	/

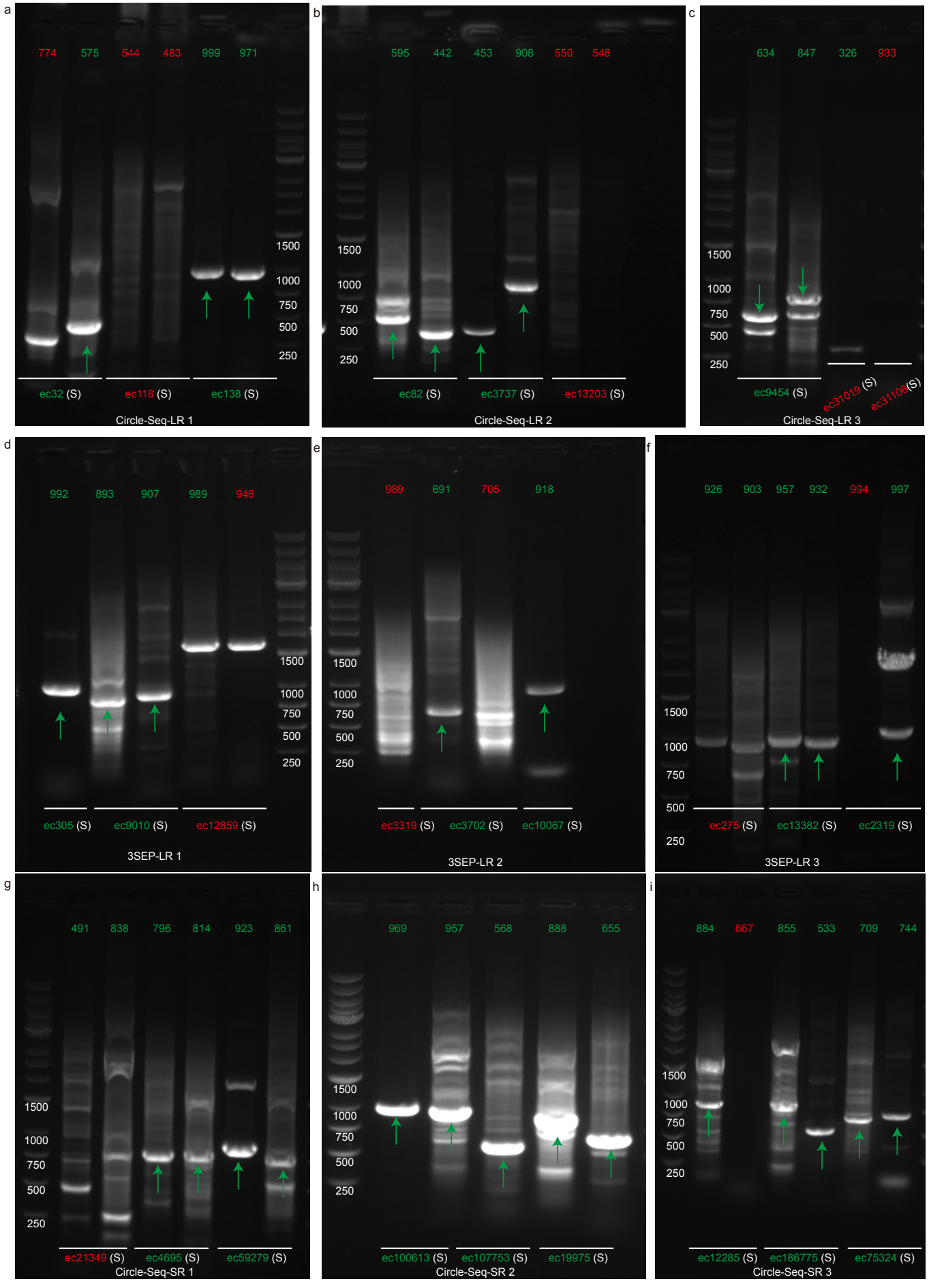
Supplementary Fig. 1. Dataset information. Column charts depicting simple eccDNA numbers and chimeric eccDNA numbers were logarithmically transformed using a base 10 logarithm.

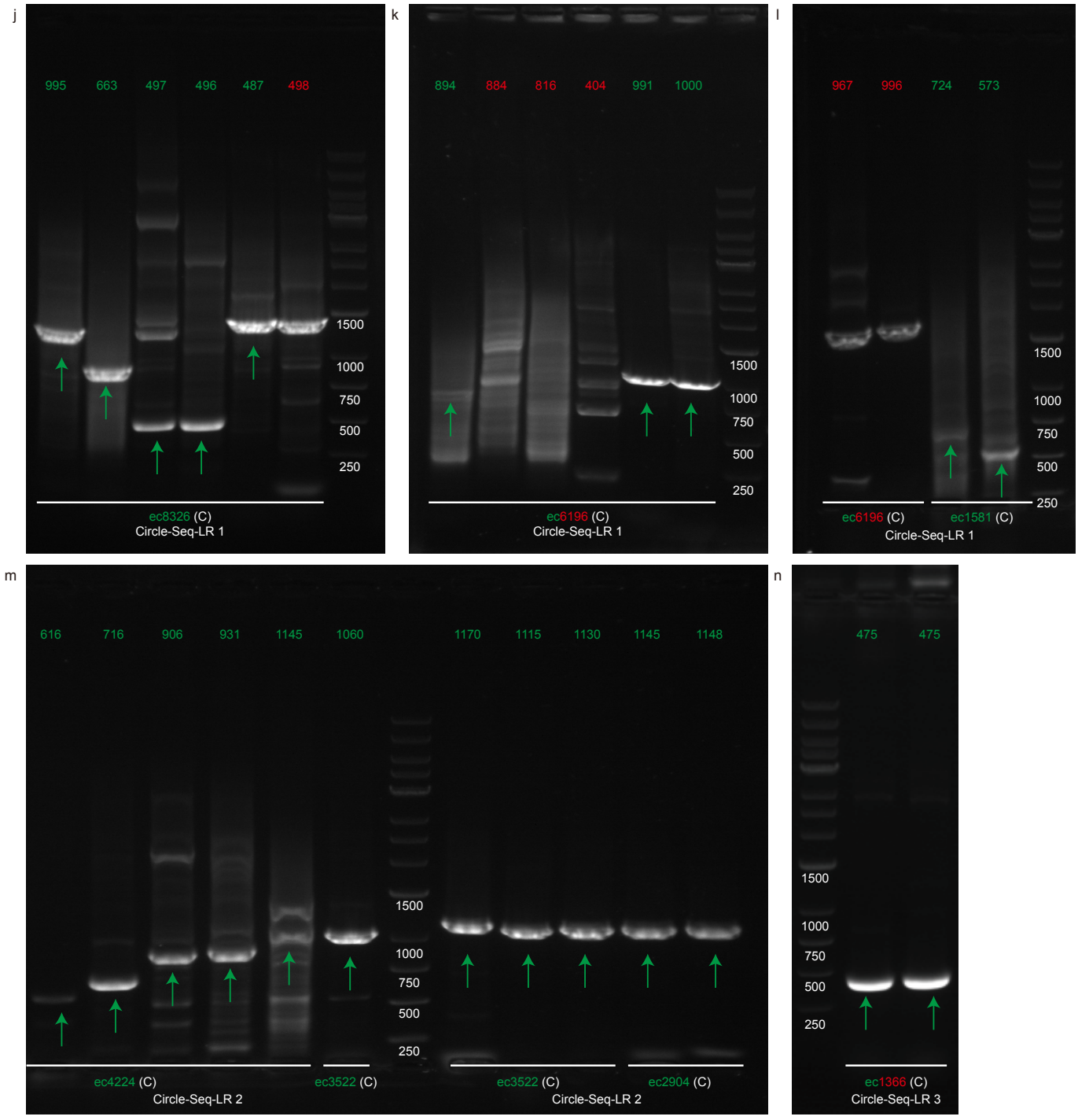


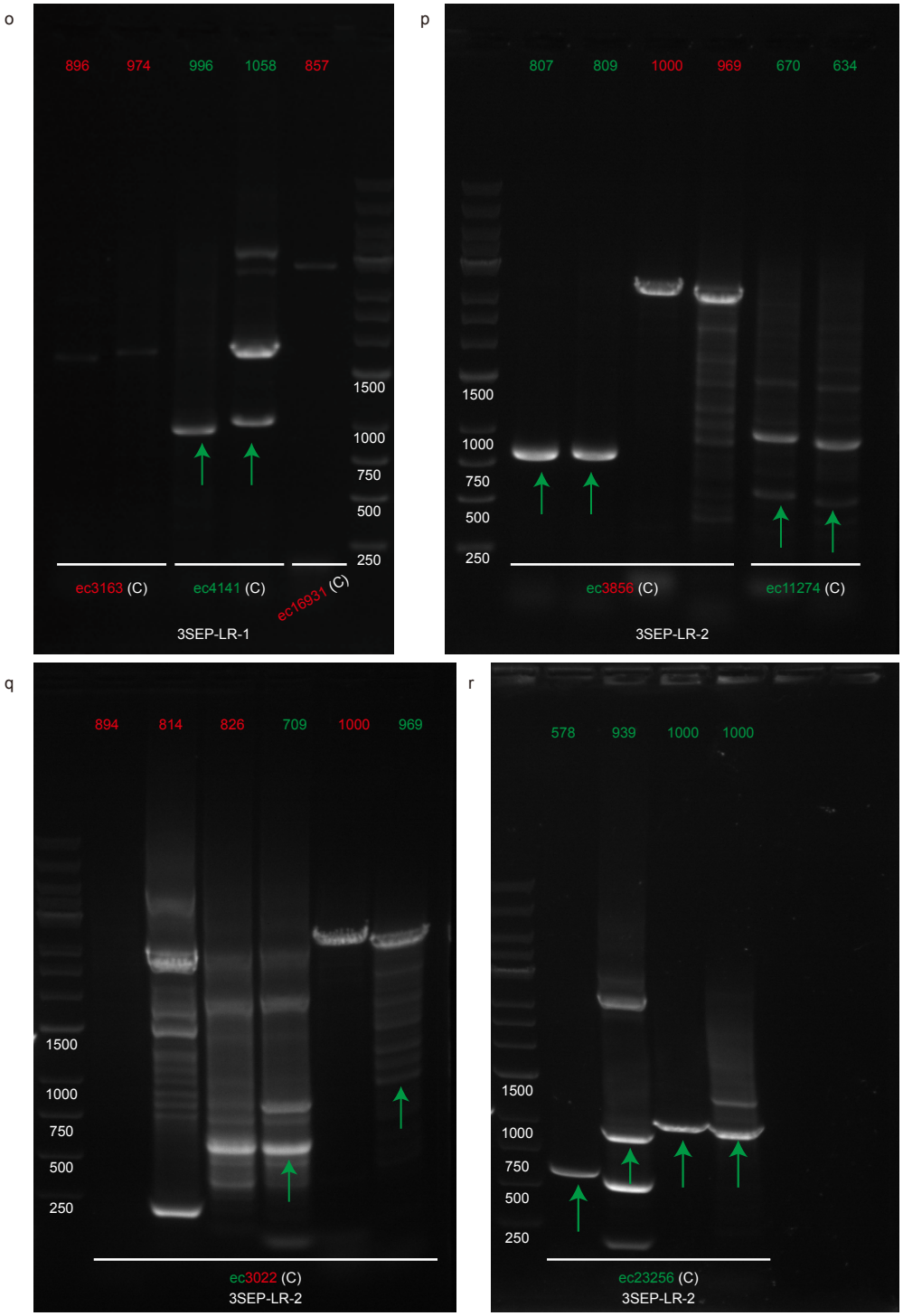
Supplementary Fig. 2. eccDNA profiles of different simulated datasets. a. Chromatin distribution comparison between the simulated dataset and corresponding real dataset. **b.** Chimeric eccDNA composition of the simulated dataset and corresponding real dataset. **c.** Length distribution comparison between the simulated dataset and corresponding real dataset.



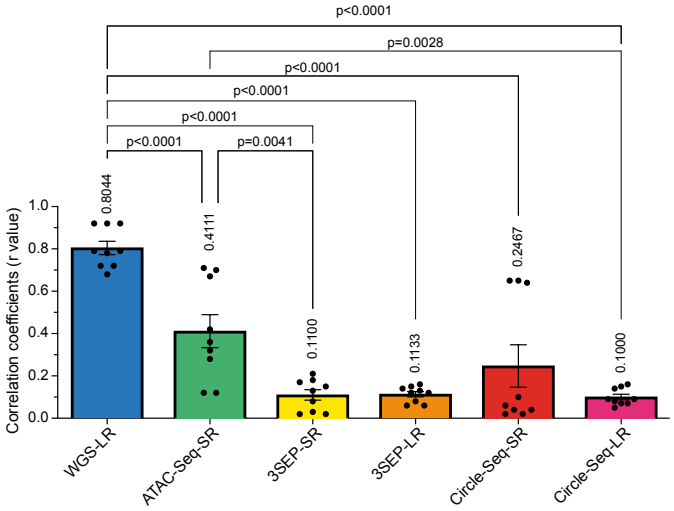
Supplementary Fig. 3. Base pair difference and computational resources cost of different analysis pipelines. a. Base pair difference of short-read data analysis pipelines under different sequencing depths. **b.** Base pair difference of long-read data analysis pipelines under different sequencing depths. **c.** Base pair difference of short-read data analysis pipelines under different chimeric DNA proportions. **d.** Base pair difference of short-read data analysis pipelines under different chimeric DNA proportions. **e.** Time and computer memory used by different pipelines in simulated short-read data analysis under different simulated sequencing depths. **f.** Time and computer memory used by different pipelines in simulated long-read data analysis under different simulated sequencing depths. The n represents the number of datasets that can be successfully analyzed by corresponding analysis pipeline. For panels **a**, **b**, **e**, **f**, n=7, except for ecc_finder (asm-sr) with n=6 and ECCsplorer with n=6 when depth ≤10, n=5 for depths between 15X and 25X, n=4 for depths between 30X and 40X, and n=3 for depths higher than 40X. For panels **c** and **d**, n=4. Data are presented as mean values +/- SEM. Source data are provided as a Source Data file.


a
774 575 544 483 999 971
1500
1000
750
500
250
ec32 (S) ec118 (S) ec138 (S)
Circle-Seq-LR 1
b
595 442 453 908 550 548
1500
1000
750
500
250
ec82 (S) ec3737 (S) ec13203 (S)
Circle-Seq-LR 2
c
634 847 326 933
1500
1000
750
500
250
ec9454 (S)
ec31019 (S)
ec31106(S)
Circle-Seq-LR 3
d
992 893 907 989 948
1500
1000
750
500
250
ec305 (S) ec9010 (S) ec12859 (S)
3SEP-LR 1
e
989 691 705 918
1500
1000
750
500
250
ec3319 (S) ec3702 (S) ec10067 (S)
3SEP-LR 2
f
926 903 957 932 994 997
1500
1000
750
500
250
ec275 (S) ec13382 (S) ec2319 (S)
3SEP-LR 3
g
491 838 796 814 923 861
1500
1000
750
500
250
ec21349 (S) ec4695 (S) ec59279 (S)
Circle-Seq-SR 1
h
969 957 568 888 655
1500
1000
750
500
250
ec100613 (S) ec107753 (S) ec19975 (S)
Circle-Seq-SR 2
i
884 667 855 533 709 744
1500
1000
750
500
250
ec12285 (S) ec186775 (S) ec75324 (S)
Circle-Seq-SR 3

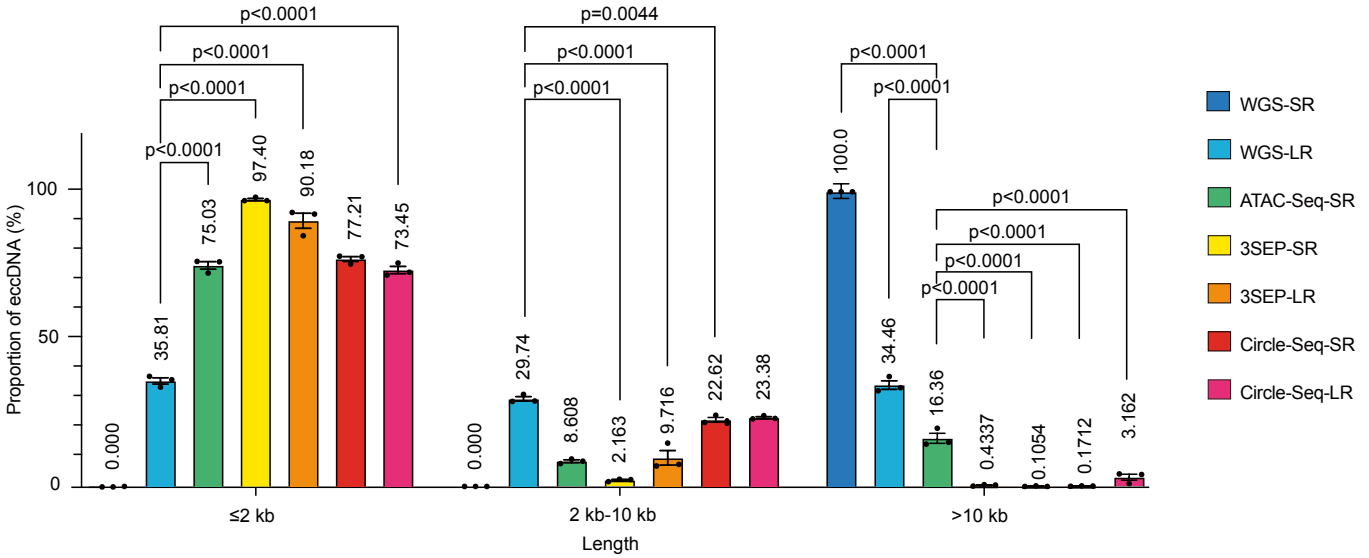

j
995
663
497
496
487
498
1500
1000
750
500
250
ec8326 (C)
Circle-Seq-LR 1
k
894
884
816
404
991
1000
1500
1000
750
500
250
ec6196 (C)
Circle-Seq-LR 1
l
967
996
724
573
1500
1000
750
500
250
ec6196 (C)
ec1581 (C)
Circle-Seq-LR 1
m
616
716
906
931
1145
1060
1170
1115
1130
1145
1148
1500
1000
750
500
250
ec4224 (C)
ec3522 (C)
Circle-Seq-LR 2
ec3522 (C)
ec2904 (C)
Circle-Seq-LR 2
n
475
475
1500
1000
750
500
250
ec1366 (C)
Circle-Seq-LR 3



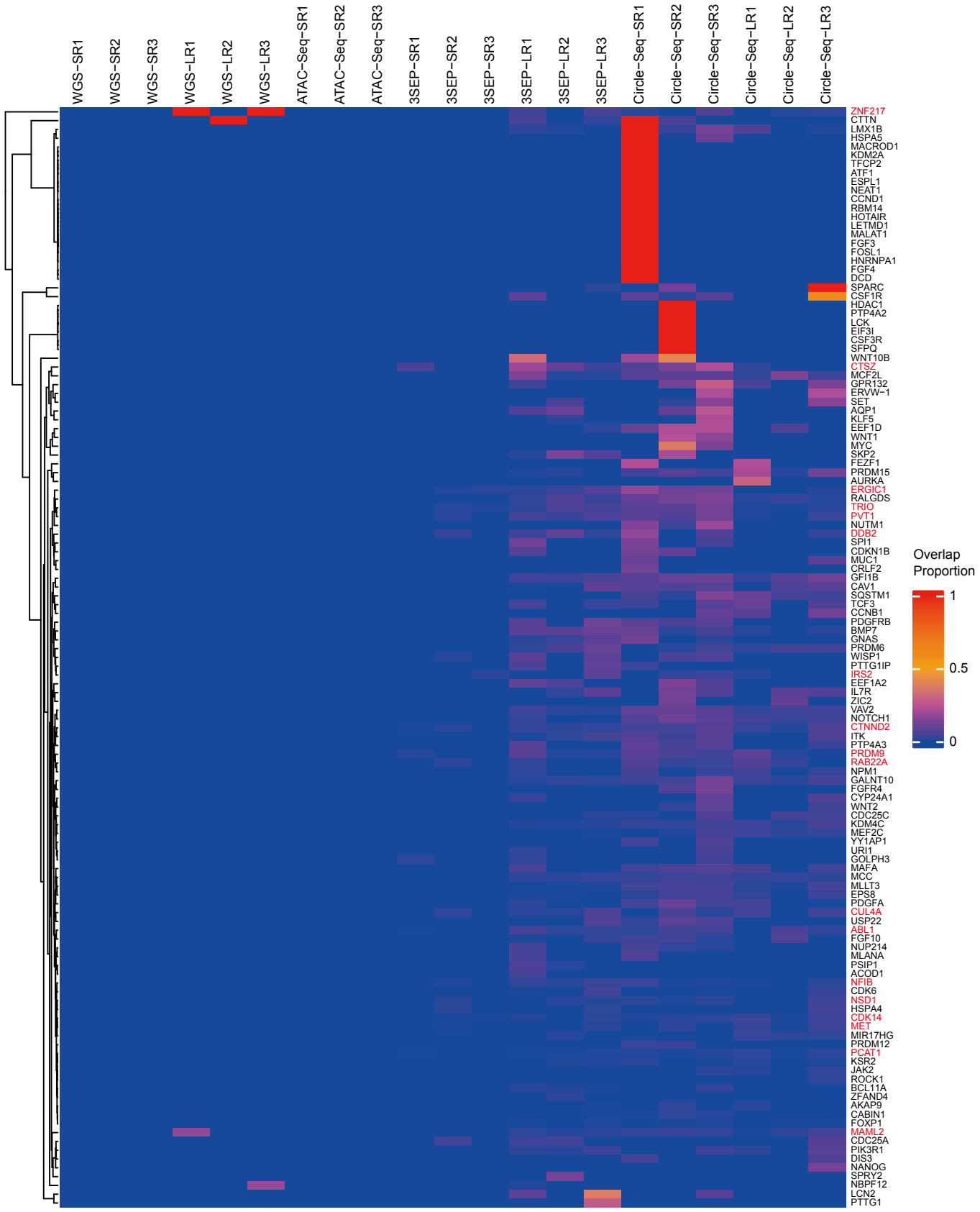
Supplementary Fig. 4. eccDNA validation by PCR. Green numbers above the lanes represented the possible amplified breakpoints and red numbers above the lanes represented the unsuccessful amplification of breakpoints. The arrow shows the validated bands. For the “ec number” below the lane, green represented the validated eccDNA, red represented the unvalidated eccDNA, and the chimeric color represented the unsuccessful for validating all the breakpoints of the specific eccDNA (only chimeric eccDNA has this label). “S” in the brackets presented the simple eccDNA, while “C” in the brackets presented the chimeric eccDNA. Green arrows showed the validated bands. The detailed Sanger sequencing validation results were recorded in Supplementary Table 2. Multiple PCR products from different samples could be run at one electrophoresis gel. To make a systematical figure layout, gels were cropped and classified in terms of the experimental method and chimeric or simple eccDNA. Panel **a**, **n**, and **b** came from the same gel from left to right; **a** and **n** shared the same ladder. Panel **o**, **c**, and **p** came from the same gel from left to right; **o** and **c** shared the same ladder. Panel **j** and **q** came from the same gel from left to right, and shared the same ladder. Panel **k** and **r** came from the same gel from left to right, and shared the same ladder. Panel **l** and **m** came from the same gel from left to right, but without shared ladder. Source data are provided as a Source Data file.



Supplementary Fig. 5. Correlation coefficients (r value) between copy number and sequencing depth of eccDNA. Copy number data were obtained from 3 WGS-LR replicates. The correlation between the eccDNA coverage from each replicate and genome copy number from each WGS-LR replicate was analyzed, resulting in a total of 9 data points for each experimental method (n=9 correlation analysis results).Statistical analysis was performed using one-way ANOVA with Tukey correction. Data are presented as mean values +/- SEM. Source data are provided as a Source Data file.



Supplementary Fig 6. Proportion of eccDNA within different length ranges. For each experimental method, we performed 3 independent experiments (n=3). Statistical analyses were performed using one-way ANOVA with Tukey correction (for each length group). Data are presented as mean values +/- SEM. Source data are provided as a Source Data file.



Supplementary Fig 7. Enriched oncogene of each replicate. Genes, marked in red, were detected by 4 experimental methods. Source data are provided as a Source Data file.

References

1. Henriksen RA*, et al.* Circular DNA in the human germline and its association with recombination. *Mol Cell* **82**, 209-217.e207 (2022).
2. Wanchai V*, et al.* CReSIL: accurate identification of extrachromosomal circular DNA from long-read sequences. *Brief Bioinform* **23**, (2022).
3. Tüns AI*, et al.* Detection and validation of circular DNA fragments using Nanopore sequencing. *Frontiers in Genetics* **13**, 867018 (2022).
4. Zhu Y*, et al.* Whole-genome sequencing of extrachromosomal circular DNA of cerebrospinal fluid of medulloblastoma. *Frontiers in Oncology* **12**, 934159 (2022).
5. Møller HD*, et al.* Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nature Communications* **9**, 1069 (2018).
6. Dillon LW*, et al.* Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell reports* **11**, 1749-1759 (2015).