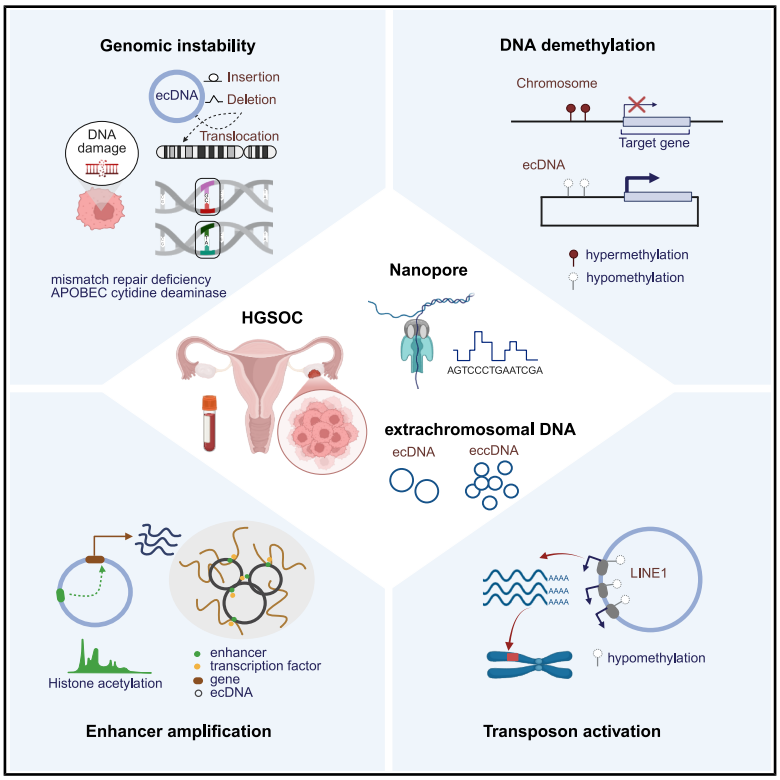
Cell Reports

Article

Landscape of extrachromosomal DNA characteristics in high-grade serous ovarian cancer via long-read sequencing

Graphical abstract



Highlights

- ecDNA suffers heavier genomic instability and lower methylation than chromosomal amplification
- ecDNA amplifies extensive enhancers, strengthening gene expression
- Transposon activation on hypomethylated ecDNA is associated with poor prognosis

Authors

Ruoming Sun, Zongkai Li, Yi Liu, ..., Ying Zhou, Kun Qu, Shouzhen Li

Correspondence

caddiezy@ustc.edu.cn (Y.Z.), qukun@ustc.edu.cn (K.Q.), lshouzhe@ustc.edu.cn (S.L.)

In brief

Sun et al. use long-read sequencing data from high-grade serous ovarian cancer to perform *de novo* profiling of extrachromosomal DNA characteristics. They show that extrachromosomal DNA (ecDNA) is a multifunctional driver of genomic variations and tumor progression, including genomic instability, enhancer amplification, and transposon activation.

Sun et al., 2025, Cell Reports *44*, 116343
October 28, 2025
https://doi.org/10.1016/j.celrep.2025.116343

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Landscape of extrachromosomal DNA characteristics in high-grade serous ovarian cancer via long-read sequencing

Ruoming Sun,[1,8] Zongkai Li,[1,8] Yi Liu,[1] Yanbing Hou,[1] Bowen Zhao,[1] Chunpeng Chen,[1] Jingwen Fang,[1,4] Chuang Guo,[1,3,7] Ying Zhou,[1,*] Kun Qu,[1,2,5,6,*] and Shouzhen Li[1,9,*]

[1]Department of Obstetrics and Gynecology, Core Facility Center, The First Affiliated Hospital of USTC, State Key Laboratory of Eye Health, 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]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, China
[4]HanGene Biotech, Hefei, China
[5]School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei, China
[6]School of Biomedical Engineering, Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, China
[7]School of Pharmacy, Bengbu Medical University, Bengbu, China
[8]These authors contributed equally
[9]Lead contact
*Correspondence: caddiezy@ustc.edu.cn (Y.Z.), qukun@ustc.edu.cn (K.Q.), lshouzhe@ustc.edu.cn (S.L.)
https://doi.org/10.1016/j.celrep.2025.116343

SUMMARY

The non-Mendelian inheritance of extrachromosomal DNA (ecDNA) exacerbates tumor genomic heterogeneity and evolution. Due to short-read sequencing's reconstruction deficiencies, ecDNA abundance in clinical tumor tissues requires further estimation, and the mechanisms driving tumor evolution remain underexplored. Here, we perform long-read whole-genome sequencing on primary and paired metastatic tumor tissues from 12 patients with high-grade serous ovarian cancer (HGSOC) and 6 normal tubal tissues, constructing a comprehensive ecDNA profile. In HGSOC, ecDNA exhibits significantly greater genomic instability and lower methylation than chromosomal DNA. Beyond oncogene and immunomodulatory gene amplification, ecDNA amplifies extensive enhancers to promote gene expression via mechanisms including enhancer hijacking and mobile enhancers. Notably, we observe activation of transposable elements on hypomethylated ecDNA, which is strongly correlated with epithelial-mesenchymal transition gene expression and poor clinical outcomes. These findings reveal ecDNA as a multifunctional driver of genomic variations and tumor progression in HGSOC, highlighting its therapeutic potential.

INTRODUCTION

High-grade serous ovarian cancer (HGSOC) is the most aggressive and common subtype of ovarian cancer, accounting for approximately 70% of all cases. HGSOC is typically diagnosed at late stages and is characterized by extensive metastasis and malignant ascites, both of which are associated with drug resistance and poor prognosis.[1] HGSOC serves as a model for studying tumor heterogeneity and genomic instability, as it is commonly associated with *TP53* mutations, homologous recombination deficiency in DNA repair, and extensive copy-number aberrations.[2] Additionally, HGSOC exhibits notable DNA hypomethylation and activation of long interspersed nucleotide element 1 (LINE1) transposons.[3,4] Circular extrachromosomal DNA, a burgeoning hallmark of human cancers, is abundant in HGSOC.[5] However, whether extrachromosomal DNA is related to these characteristics of HGSOC remains unclear, and the mechanisms by which it drives HGSOC progression have not yet been investigated.

Extrachromosomal DNA is defined as circular, acentric chromatin molecules. Based on copy number and size, extrachromosomal DNA is mainly divided into two categories: one is large and copy number-amplified circular extrachromosomal DNA (ecDNA), which has recently been recognized as a major contributor to tumor evolution,[5,6] and the other is small extrachromosomal circular DNA (eccDNA), which exists in both tumor and normal cells.[7,8] ecDNA possesses unique characteristics that make it a potent driver of tumor progression. Because ecDNA lacks centromeres, it is unevenly distributed into daughter cells during mitosis, rapidly increasing genomic heterogeneity within the cell population. Moreover, ecDNA can extensively amplify oncogenes and contains highly accessible chromatin, resulting in elevated expression of these amplified genes, which provides tumor cells with a significant selective advantage.[9–11]

Additionally, ecDNA can aggregate into hub-like structures, enabling complex transcriptional regulatory interactions. These interactions promote high expression of cancer-associated genes through mechanisms such as enhancer hijacking and mobile enhancers.[12–14] Furthermore, coordinated inheritance of ecDNA species can quickly establish selective advantages within tumors.[15,16] APOBEC3 mutagenesis of ecDNA and ecDNA rearrangement indicate its intrinsic evolutionary capacity.[17–19] Combined with its ability to reintegrate into the genome,[20] ecDNA further exacerbates genomic heterogeneity in tumors. Nevertheless, whether ecDNA drives tumor progression through other mechanisms urgently needs further research.

The accumulation of public next-generation sequencing (NGS) data has enabled the detection of ecDNA in a wide range of tumor types using short-read sequencing.[5,6,17] However, due to the limitations of short-read sequencing in reconstructing complex genomic regions, the abundance and structure of ecDNA in clinical samples have not been effectively evaluated. Long-read sequencing, with its capability of producing reads up to 100 times longer than those from NGS platforms, offers a significant advantage for assembling complex genomic regions such as repetitive sequences and transposable elements (TEs). This technology facilitates the identification of broad genetic variations and chromosomal rearrangements, providing an opportunity to construct complete ecDNA structures. Another notable feature of long-read sequencing is its ability to directly detect DNA methylation in native DNA without requiring bisulfite conversion, enabling the simultaneous assessment of genomic element activity.[21,22] While long-read sequencing has been applied to validate ecDNA structures,[15,23–26] its use in the *de novo* identification and comprehensive profiling of ecDNA in clinical samples remains unexplored.

The advent of new sequencing technologies presents an unprecedented opportunity to dissect the oncogenic mechanisms of ecDNA with increasing precision. In this study, we systematically profiled the ecDNA characteristics in HGSOC using long-read whole-genome sequencing (WGS), with normal fallopian tube (FT) tissues as a control. This approach allowed us to explore the potential oncogenic mechanisms driven by ecDNA in HGSOC. By leveraging these advanced sequencing technologies, our findings offer valuable insights into the role of ecDNA in driving tumor evolution and open new avenues for targeted therapeutic interventions in cancer treatment.

RESULTS

Long-read WGS identifies abundant ecDNAs and eccDNAs in HGSOC

To elucidate the characteristics of ecDNA in HGSOC, we collected primary and paired metastatic tumor tissue samples from 12 patients with HGSOC, along with normal FT tissues from 6 non-tumor donors as controls. WGS using Oxford Nanopore Technologies (ONT) was performed on all 30 samples (Figure 1A; Table S1). Additionally, RNA sequencing (RNA-seq) was conducted on 21 tumor tissue samples and 6 normal FT tissues. WGS using Illumina was carried out on peripheral blood mononuclear cell (PBMC) samples matched to 20 tumor tissues as a reference for germline mutations. To minimize variability caused by treatment, only treatment-naive patient samples were included in this study.

After filtering, we obtained 30 high-quality ONT WGS datasets, with a median genome coverage of 24.86× and a read N50 of 25.5 kb (Figures S1A and S1B; Table S1). Highly consistent with published HGSOC copy number variation (CNV) profiles from TCGA,[27] our CNV analysis of ONT WGS data using QDNAseq software revealed extensive copy-number gains and losses across entire chromosomal arms in HGSOC tumor tissues,[28] whereas no CNVs were detected in FT samples (Figure S1C). We used CReSIL, a computational tool proven to effectively identify extrachromosomal DNA from long-read WGS data, to detect eccDNA in the samples[29,30] and identify ecDNA together with Decoil.[31] Excitingly, a total of 3,833 extrachromosomal DNAs were identified across the 30 samples, with a total of 3,133 eccDNAs (range from 44 to 296) in HGSOC tissues compared to 582 eccDNAs (range from 61 to 133) in normal FT tissues and a total of 117 ecDNAs (range from 0 to 17) in HGSOC tissues compared to only one ecDNA in normal FT tissues (Figure 1B; Table S2). As expected, our data showed a positive correlation between the number of genes per megabase and eccDNA abundance, consistent with prior studies (Figure S2A).[7,25]

We categorized the identified circular DNAs into simple circular DNAs and complex circular DNAs based on circular junctions forming from a single continuous locus or multiple genomic fragments originating from the same or different chromosomes (Figures S2B and S2C). Compared to eccDNA, ecDNA had a higher proportion of complex types (Figure 1B). Using the ONT long-read WGS method and ecDNA detection algorithms, we achieved comprehensive profiling of each ecDNA, including complete sequences, gene annotations, mutations, and DNA methylation profiles, enabling further exploration of their oncogenic properties. For instance, we found that an ecDNA from paired OC8P and OC8M samples was assembled from a chromosomal fragment on chr11 and contained the complete *DDB2* gene, a known oncogene amplified in ovarian cancer (Figure 1C).[5,32] To validate the amplification and circular structure of ecDNA, we performed interphase DNA fluorescence *in situ* hybridization (FISH) and Circle-seq (a highly sensitive circular DNA purification method) on the matched frozen HGSOC samples. DNA FISH images showed that many oncogene spots were scattered in the cell nucleus far away from the corresponding centromeres, supporting extrachromosomal amplification (Figures 1D and S2D). Meanwhile, Circle-seq also verified the circular closed structure of ecDNAs containing *DDB2* and *MDM2* in HGSOC tissues (Figure S2E). Furthermore, *de novo* sequence assembly of the long reads mapping to the oncogene-amplified regions allowed us to further physically reconstruct their circular structure (Figure S2F).

Genomic instability in circular DNA

Genomic instability is known to play an important role in tumor evolution by increasing genomic heterogeneity, and previous studies have demonstrated that genomic instability is widely associated with ecDNA production.[15,20] In this study, we analyzed genomic instability in ecDNA and eccDNA from HGSOC using long-read WGS data, focusing on structural

CellPress
OPEN ACCESS

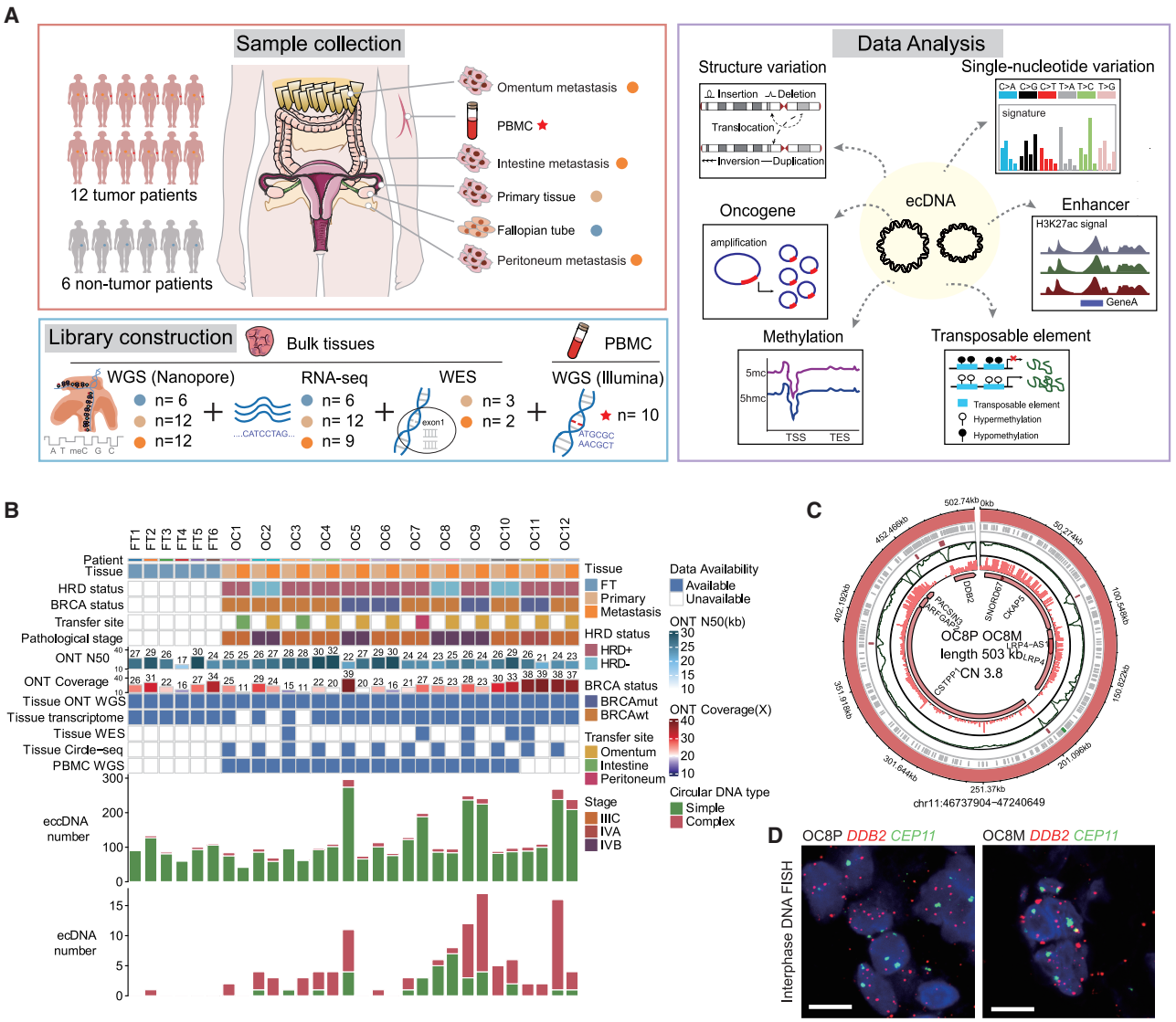


Figure 1. The workflow and identification of circular DNA from whole-genome ONT sequencing

(A) The workflow diagram illustrates the sampling and analysis strategies.

(B) Distinct circular DNA profile of 30 samples. The number of identified ecDNAs and eccDNAs for each sample is indicated in the bar plot. Patients' condition, clinicopathological information, and sequencing technology are color coded. Simple circular DNA, contains a single locus; complex circular DNA, contains multiple regions from the same chromosome or different chromosomes.

(C) Circos shows an example of ecDNA annotation (OC8P and OC8M). The circle annotations from outside to inside are transposable element (LINE1, AluY, and LTR), enhancer, DNA methylation, gene expression, and gene.

(D) Representative DNA FISH images of interphase ecDNA from matched frozen HGSOC tissues (OC8P and OC8M) showing extrachromosomal *DDB2* signals against centromere labeling (CEP11) signals. Scale bars, 10 μm.

See also Figures S1 and S2 and Tables S1 and S2.

variation (SV) and single-nucleotide variation (SNV), as distinct from chromosomal DNA.

The excellent alignment capability of long-read sequencing can efficiently identify SVs. Using Sniffles2,[33] we identified approximately 24,859 SVs and 7,890 mosaic SVs (low frequency, 2%–30% variant allele frequency) per sample (Table S3). SV length distribution exhibited distinct peaks at ~300 bp and ~6 kb, consistent with the lengths of ALU (a type of short interspersed nuclear element, SINE) and LINE transposons, respectively (Figure S3A), demonstrating the sensitivity and accuracy of SV identification, and 20 of 55 genomic loci with recurrent SV breakpoints were near HGSOC-related oncogenes,[19] such as *KRAS*, *RHO*, and *ARID3A* (Figure S3B; Table S3). We assessed SV frequencies in the chromosomal genome, chromosomal amplification, and ecDNA regions. Surprisingly, ecDNA showed significantly higher SV frequencies than genome and

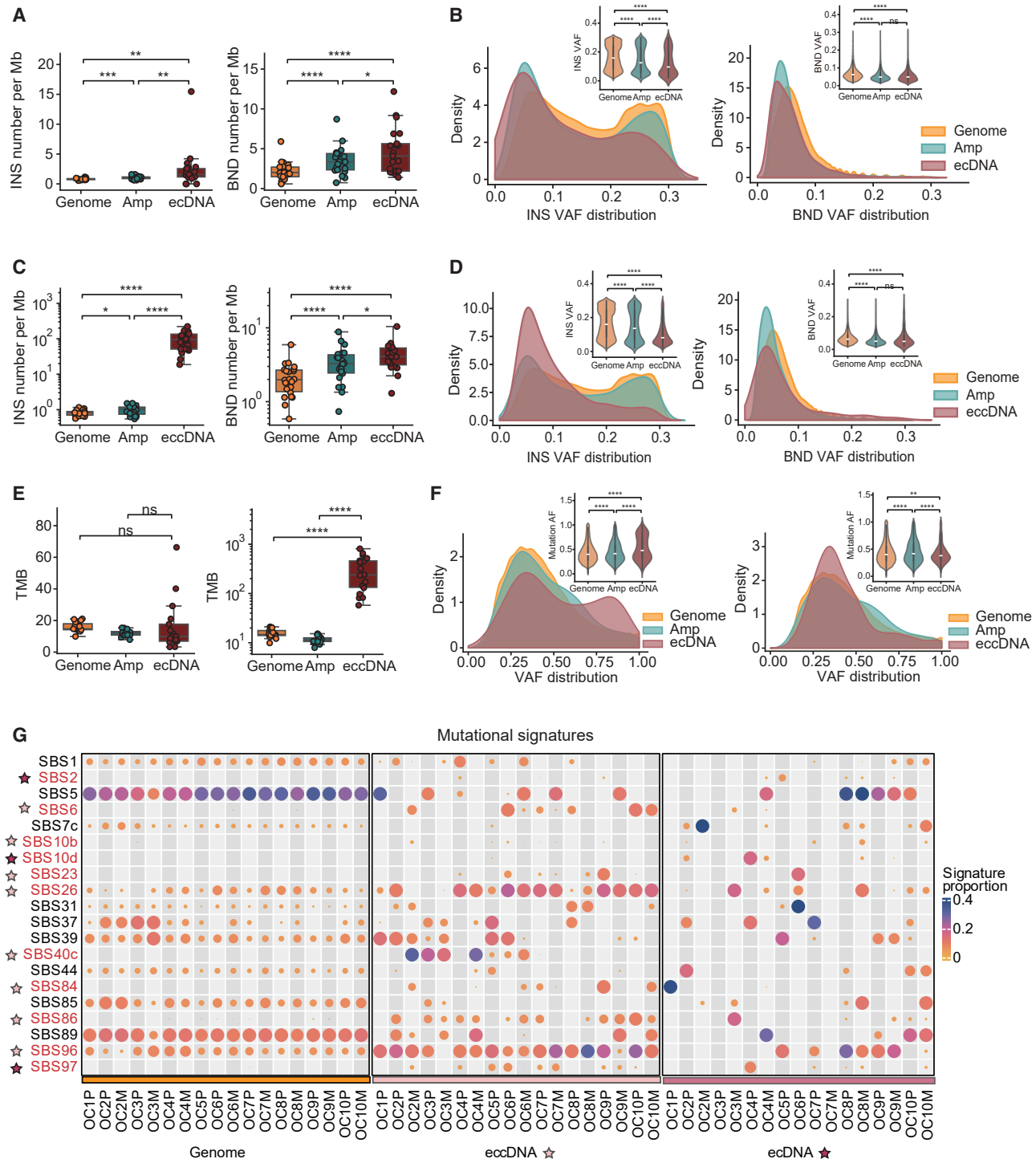


Figure 2. Genomic instability of circular DNA in HGSOC

(A) Comparison of the SV number per Mb detected in ecDNA, whole genomic regions, and amplified regions in each tumor sample (two-sided Wilcoxon's signed rank-sum test, $n = 22$; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$).

(B) Comparison of the SV VAF distribution detected in ecDNA, whole genomic regions, and amplified regions in all tumor samples (Mann-Whitney U test, $n = 22$; $^{****}p < 0.0001$).

(C) Comparison of the SV number per Mb detected in eccDNA, whole genomic regions, and amplified regions in each tumor sample (two-sided Wilcoxon's signed rank-sum test, $n = 22$; $^{*}p < 0.05$ and $^{****}p < 0.0001$).

(legend continued on next page)

chromosomal amplification (Figures 2A and S3C), in terms of insertions ($p < 0.0001$, paired Student's t test) and breakends ($p < 0.0001$). Furthermore, compared to chromosomal genomes, the SVs of ecDNA displayed lower allelic variant frequencies (VAFs) ($p < 0.0001$, Mann-Whitney U test) (Figures 2B and S3D), suggesting that these SVs may occur later in the ecDNA life cycle. Notably, these characteristics were conserved in eccDNA (Figures 2C, 2D, S3C, and S3D), suggesting that they are a feature of extrachromosomal DNA. The presence of tree-rearrangement clusters overlapping with ecDNAs further supported the hypothesis that ecDNA served as a substrate for genomic remodeling,[20] contributing to the aberrant expression of tumor-suppressor genes and oncogenes (Figures S3E and S3F).

To evaluate SNVs in HGSOC ecDNA, we used Clair3[34] and identified a total of 7,915 somatic functional SNVs of 18 tumor samples after filtering from long-read sequencing (Table S3), whose precision reached 98.9% by comparing to Illumina-based whole-exome sequencing (WES) (Figure S3G). Many SNVs affected frequently mutated genes in HGSOC, such as *TP53*, which is associated with poor prognosis, and *MUC16*, which is expressed and secreted in HGSOC ascites tumor cells (Table S3).[35] When comparing the tumor mutational burden (TMB), the total number of SNVs per megabase, eccDNA showed a significantly higher TMB than chromosomal genome and amplification ($p < 0.0001$, paired Student's t test) (Figure 2E), while there was no difference for ecDNA, consistent with the WES results (Figure S3H). Additionally, the SNVs of eccDNA exhibited lower VAFs compared to chromosomal regions ($p < 0.0001$, Mann-Whitney U test) (Figure 2F), indicating a higher rate of acquired SNVs. However, ecDNA had significantly higher VAFs compared to chromosomal regions ($p < 0.0001$, Mann-Whitney U test) (Figure 2F), which might suggest positive selection of these SNVs, consistent with a previous study.[17] Furthermore, we performed COSMIC single-base substitution (SBS) signature analysis based on a Bayesian non-negative matrix factorization (NMF) algorithm (Figure 2G; Table S3) to explore the broad spectrum of mutational processes of eccDNA and ecDNA.[36,37] The result revealed that signatures of mismatch repair deficiency (SBS6 and SBS26) and polymerase epsilon exonuclease domain mutations (SBS10b) were significantly enriched in eccDNAs compared with the chromosomal genome ($p < 0.05$, Wilcoxon rank-sum test) (Figures 2G and S3J), while ecDNAs highly enriched the signatures of APOBEC cytidine deaminase (SBS2) and defective POLD1 (DNA polymerase δ 1) proofreading (SBS10d) compared with the chromosomal genome ($p < 0.05$, Wilcoxon rank-sum test) (Figures 2G and S3I). In contrast, SNVs in the chromosomal genome were enriched for clock-like signatures (SBS1 and SBS5) ($p < 0.01$; Wilcoxon rank-sum test). These findings suggest that ecDNA and eccDNA in HGSOC are more prone to defects in DNA repair, exacerbating the TMB.

In summary, compared to linear chromosomal DNA, ecDNA and eccDNA in HGSOC exhibited greater genomic instability. This instability may enhance tumor genomic heterogeneity, working in concert with the uncoordinated inheritance mode of ecDNA, and provide a driving force for tumor evolution.

Distinct extrachromosomal DNA characterization between HGSOC and FTs

To investigate the different characteristics in circular DNA between normal FTs and HGSOC tumor tissues, we compared eccDNA and ecDNA in terms of number, size, structural complexity, genomic origin, and DNA methylation. As expected, we observed that the number of circular DNA in HGSOC tumor tissues was significantly higher than that in normal FT tissues, regardless of eccDNA and ecDNA ($p = 0.039$ and 4.31e−05, respectively; independent Student's t test) and no matter whether the sequencing depth was normalized or not (Figures S4A–S4C). Furthermore, the average length of ecDNA in HGSOC was significantly longer than normal FT ($p = 0.0017$, independent Student's t test), as well as eccDNA ($p = 0.00055$) (Figure S4D). Additionally, the eccDNA and ecDNA of HGSOC exhibited greater structural complexity, evidenced by higher fragment numbers than those in normal FT ($p = 0.0015$ and 0.00063, respectively; independent Student's t test) (Figure S4E). Unsurprisingly, no significant difference in the copy number of eccDNA was observed between HGSOC and FT, which made the amplification feature of ecDNA prominent in HGSOC ($p = 0.0015$, independent Student's t test) (Figure S4F). To explore the genomic origins of circular DNAs, we applied locus overlap analysis (LOLA) for genomic region enrichment analysis[38] and found that circular DNAs in HGSOC tissues were significantly enriched in transcription start sites (TSSs), transcribed regions, enhancers, and simple repeat segments, while circular DNAs in normal FT were enriched in genomic super duplications (two-sided Fisher's exact test, $p < 0.05$) (Figure S4G), highlighting the gene-encoding and regulatory functionality of circular DNAs in HGSOC. However, there was no difference in circular DNA between primary and metastatic tumors in terms of these characteristics.

The accurate DNA methylation (5-methylcytosine [5mC]) and hydroxymethylation (5-hydroxymethylcytosine [5hmC]) data were retained from ONT sequencing (Figures S5A and S5B),[39]

(D) Comparison of the SV VAF distribution detected in eccDNA, whole genomic regions, and amplified regions in all tumor samples (Mann-Whitney U test, $n = 22$; ****$p < 0.0001$). INS, insertion; BND, translocation.

(E) Comparison of the somatic SNV TMB detected between ecDNA (left) and eccDNA (right), whole genomic regions, and amplified regions in each tumor sample (two-sided Wilcoxon's signed rank-sum test, $n = 18$; ****$p < 0.0001$). TMB, tumor mutational burden.

(F) Comparison of the somatic SNV VAF distribution detected between ecDNA (left) and eccDNA (right), whole genomic regions, and amplified regions (Mann-Whitney U test, $n = 18$; **$p < 0.01$ and ****$p < 0.0001$). VAF, variant allele frequency.

(G) SBS signature enrichment scores are shown by both dot size and heat scale. Increasing dot sizes indicate increasing enrichment of indicated SBS signatures within genomic, eccDNA, and ecDNA regions. The circular DNA SBS signatures with significant enrichment compared with both genomic and amplified regions are marked with five-pointed stars beside them (Wilcoxon rank-sum test, $p < 0.05$). Pink represents eccDNA enrichment, and purple represents ecDNA enrichment. The rest of the SBS signatures are enriched in the genome. Excluded from the analysis are tumors without patient-matched PBMCs.

Unless otherwise specified, boxes show the median and the interquartile range (IQR); the lower whisker indicates Q1 – 1.5× the IQR; the upper whisker indicates Q3 + 1.5× the IQR. See also Figure S3 and Table S3.

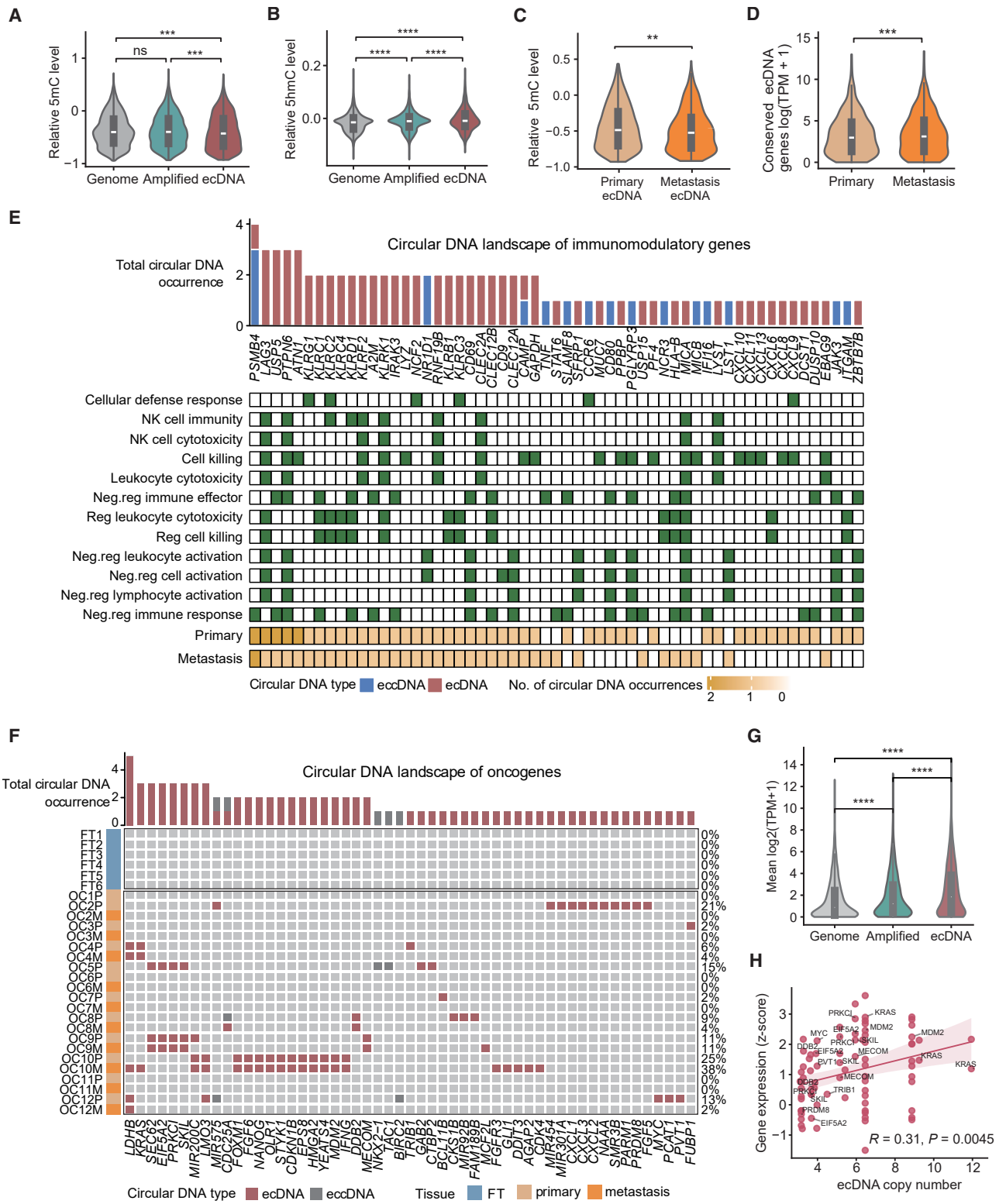


Figure 3. Demethylation and gene amplification characteristics of ecDNAs in HGSOC

(A and B) Relative 5mC (A) and 5hmC (B) levels between genomic, amplified, and ecDNA regions for HGSOC samples (two-sided independent Student's *t* test; ***$p < 0.001$ and ****$p < 0.0001$).

(C) Relative 5mC levels in ecDNA regions for primary and metastatic samples (two-sided independent Student's *t* test; **$p < 0.01$).

(legend continued on next page)

which allowed us to uncover the epigenetic characteristic of ecDNA. Unsurprisingly, compared to normal FT, the genome of HGSOC showed much lower 5mC levels and modestly lower 5mC levels for circular DNA (Figure S5C). In detail, we compared the differential methylation of the genome, chromosomal amplification, and ecDNA regions. Interestingly, the methylation level of ecDNA was significantly lower than that of genome and chromosomal amplification (p = 0.00087 and 0.0009, respectively; independent Student's t test) (Figure 3A), suggesting that the epigenetic change of ecDNA may enhance the activity of regulatory elements,[9,40,41] altering gene expression patterns. Furthermore, the 5hmC level of ecDNA was significantly higher than that of genome and chromosomal amplification (p = 5.43e−46 and 1.58e−15, respectively; independent Student's t test) (Figure 3B), indicating that ecDNA might undergo active DNA demethylation, potentially facilitating the functional activation of these extrachromosomal elements. Notably, these phenomena were absent in eccDNA (Figure S5D), highlighting the distinct characteristic of ecDNA.

In 12 patients with paired primary and metastatic tumor samples, an average of 52.3% of circular DNAs (21.5% for ecDNA and 52.8% for eccDNA) were shared (>80% sequence overlap) between paired samples, reflecting the persistence and clonal selection of circular DNA to some extent (Figure S5E). To reveal the evolutionary trend of ecDNA, we compared the methylation of ecDNAs between primary and metastatic tumors. Surprisingly, the ecDNAs of metastasis displayed significantly lower 5mC levels (p = 0.0012, independent Student's t test) (Figures 3C and S5F), which might represent the enhancing activity of gene expression in metastasis. Actually, the conserved genes amplified in ecDNAs of paired tumoral samples indeed were highly expressed in metastatic tumors (p = 0.00014, paired Student's t test) (Figure 3D). Besides, this epigenetic change did not occur in eccDNA.

Taken together, these results reveal that HGSOC tumor tissues are enriched with functional and activated circular DNAs, in stark contrast to normal tissues. Beyond this, compared to chromosomal amplification, ecDNA exhibited a distinct demethylation characteristic, and the lower methylation along with more active transcription may represent the evolutionary trend of ecDNAs in HGSOC.

HGSOC ecDNAs amplify immunomodulatory genes and oncogenes

As the genes amplified in ecDNAs varied among individuals, to unveil the conserved characteristics of gene amplification, we performed gene set enrichment analysis (GSEA) to examine the major pathways associated with ecDNAs. In HGSOC, both ecDNA and eccDNA showed significant enrichment in immunomodulatory processes, such as cell killing (p = 2.21e−07; two-sided Fisher's exact test), natural killer cell-mediated immunity (p = 3.01e−05), adaptive immune response (p = 2.23e−05), and regulation of chronic inflammatory response (p = 7.07e−05), involving genes like *PSMB4*, *LAG3*, *USP5*, *PTPN6*, and so on (Figures 3E and S5G). This suggests that ecDNA and eccDNA play a role in interactions between the tumor and immune microenvironment. In contrast, for the eccDNA of normal FT, a biological process related to immunoglobulin-mediated immune response was significantly enriched (p = 1.10e−07; two-sided Fisher's exact test) (Figure S5G), consistent with the presence of circular DNA as a by-product of V(D)J recombination in normal tissues.[7,42,43]

In the meantime, our data revealed that circular DNAs harboring intact oncogenes exclusively existed in HGSOC tissues, especially dominant in ecDNAs (Figure 3F). Furthermore, there were many oncogene-amplified ecDNAs conserved between primary and metastatic tumors, such as *LDHB*, *KRAS*, *MDM2*, *DDB2*, and so on. This characteristic reflected the hereditary and selective advantages of ecDNAs. Eleven ecDNAs were found to include multiple oncogenes in a single event. For instance, an ecDNA derived from contiguous genomic regions on chromosome 3q26 contained the *SKIL* gene and adjacent oncogenes *PRKCI* and *EIF5A2* (Figure S5H). *EIF5A2* is a frequently amplified oncogene previously reported in primary ovarian cancer.[44]

The current view holds that ecDNAs establish a selective advantage by increasing the copy number of the amplified gene.[45] As expected, compared to genome and chromosomal amplification, genes amplified on ecDNAs exhibited significantly higher expression (p = 8.1e−30 and 1.2e−11, respectively; independent Student's t test) (Figure 3G), with no difference between chromosomal amplification and eccDNA (Figure S5I). In particular, for oncogene-amplified ecDNAs, there was a significant positive correlation between the copy number of ecDNAs and transcription levels (R = 0.31, p = 0.0045; Pearson's correlation) (Figure 3H). Intriguingly, some ecDNA genes with low copy numbers exhibited high expression levels, suggesting the involvement of additional regulatory mechanisms, such as enhancers and other transcriptional regulatory elements, in enhancing ecDNA gene transcription.

Overall, our data indicate that in HGSOC, ecDNAs mainly amplified immunomodulatory genes and oncogenes to promote tumor progression. However, in addition to copy amplification, there may be other regulatory mechanisms enhancing the expression of genes in ecDNAs.

Extrachromosomal DNAs amplify extensive enhancers, strengthening gene expression

The junction and circularization of chromosomal fragments in circular DNAs result in extensive genomic rearrangements, leading

(D) Gene expression of the conserved ecDNA derived from paired primary and metastatic tumor tissues from the same patients with HGSOC (two-sided paired Student's t test; ***$p < 0.001$).
(E) Top: bar plot showing the total number of occurrences (y axis) of immunomodulatory genes on circular DNA (x axis). Middle: heatmap showing Gene Ontology (GO) terms associated with immune genes (green). Bottom: tissue types in which the immune genes are observed to be on circular DNA.
(F) Oncoplot shows oncogenes distribution on ecDNA and eccDNA of all samples.
(G) Gene expression comparing located on genomic, amplified, and ecDNA regions for HGSOC samples (two-sided independent Student's t test; ****$p < 0.0001$).
(H) Pearson's correlation analysis of gene expression and copy number of oncogenes amplified on HGSOC ecDNAs (R = 0.31, p = 0.0045, n = 81 genes).
See also Figures S4 and S5.

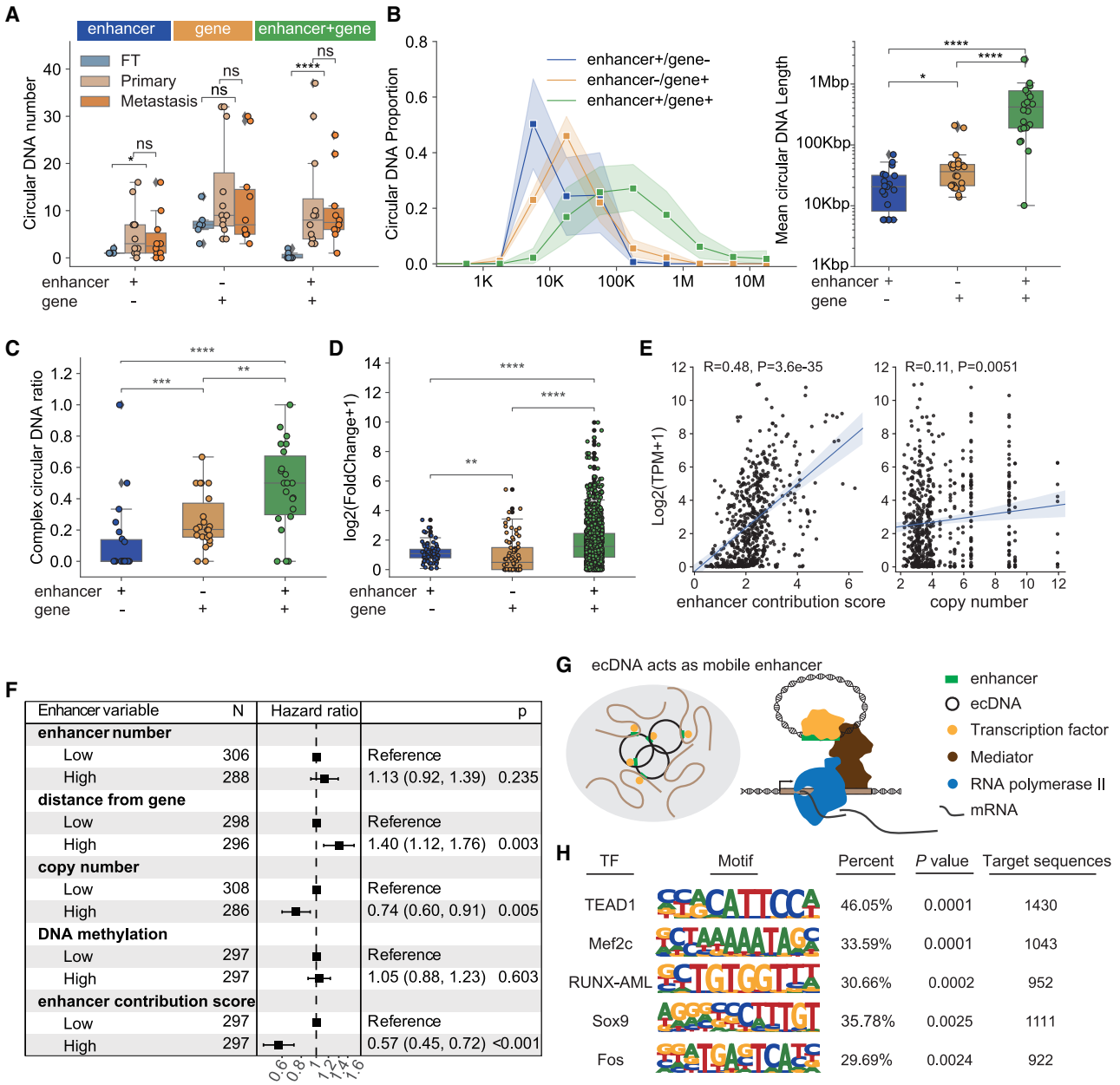


Figure 4. Diverse enhancers on circular DNA promote gene expression

(A) Comparison of the number of circular DNA with enhancers, genes, and co-amplification for FT, primary, and metastatic HGSOC samples (Brunner-Munzel test and paired Student's *t* test; $*p < 0.05$ and $****p < 0.0001$).

(B) Comparison of the length distribution of circular DNA with enhancers, genes, and co-amplification for HGSOC samples (Mann-Whitney U test; $*p < 0.05$ and $****p < 0.0001$).

(C) Comparison of the ratio of complex circular DNA with enhancers, genes, and co-amplification for HGSOC samples (Mann-Whitney U test; $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$).

(D) Comparison of the coding-gene expression of circular DNA with enhancers, genes, and co-amplification (Mann-Whitney U test; $**p < 0.01$ and $****p < 0.0001$).

(E) Pearson's correlation of enhancer contribution score and gene expression of coding genes on enhancer+/gene+ circular DNA vs. normalized coverage. Point estimates of two-sided Pearson correlation coefficient test and their 95% confidence level intervals (in blue) are shown.

(F) Forest plots of multivariable gene expression Cox regression analysis for enhancer+/gene+ circular DNA subtypes with enhancer elements variables as a confounding factor in HGSOC tumor tissues. The point estimations of the hazard ratio derived from the Cox regression test and their corresponding 95% confidence level intervals (error bars) are presented. Here, the median as a cutoff is adopted for classifying these enhancer element variables into high and low groups.

(legend continued on next page)

to rewired gene regulatory forms, such as enhancer hijacking.[12–14] Next, we further explored the impact of enhancers amplified on circular DNA. Utilizing publicly available H3K27ac chromatin immunoprecipitation (ChIP)-seq data from HGSOC tumor samples,[46] we identified 30,306 enhancers, which showed strong anti-correlation with DNA methylation status from ONT data (Figures S6A and S6B), validating data consistency between the two sources. 356 out of 3,715 circular DNAs in all samples (n = 28) contained intact enhancers, with a total of 3,227 enhancers amplified (Table S4).

To assess the functional impact of enhancer-amplified ecDNA in HGSOC, we classified circular DNAs into three categories: (1) only enhancer amplified (enhancer+/gene–), (2) only gene amplified (enhancer–/gene+), and (3) gene and enhancer co-amplified (enhancer+/gene+). The proportion of circular DNA carrying intact enhancers was significantly higher in HGSOC than in normal FT (p = 1.71e−02 and p = 5.46e−22; Brunner-Munzel test), while there was no significant difference in circular DNA carrying only intact genes (Figure 4A). Logically, enhancer +/gene+ circular DNAs exhibited larger sizes (p < 0.0001; Mann-Whitney U test) (Figures 4B and S6C) and greater structural complexity (p = 5.10e−05 and p = 4.90e−03; Mann-Whitney U test) (Figure 4C), which matched ecDNA features and suggested potential rewired gene regulation, respectively. Combined with RNA-seq data, we found that genes amplified on enhancer+/gene+ ecDNAs had significantly higher expression levels than those in other categories (p < 0.0001; Mann-Whitney U test) (Figure 4D), reflecting the importance of a complete gene regulatory framework. Furthermore, enhancer-regulated genes were enriched in pathways such as AKT signaling, regulation of leukocyte-mediated immunity, cell division, and RHO GTPase cycle (Figures S6D and S6E), which are closely associated with HGSOC progression.[4]

Interestingly, gene expression for all circular DNA-amplified coding genes showed a modestly positive correlation with the copy number (R = 0.11, p = 0.0051; Pearson's correlation) (Figure 4E). To systematically uncover the effect of enhancers on circular DNA to gene expression, we constructed a ridge regression model defining an enhancer contribution score based on several key properties: number, distance from the gene, copy number, and DNA methylation of circular DNA enhancers. Compared to the mild correlation of ecDNA copy number, we observed a significant positive correlation between gene expression and enhancer contribution scores (R = 0.48, p = 3.6e−35; Pearson's correlation) (Figure 4E). Multivariable Cox regression analysis confirmed that the enhancer contribution score was a statistically significant factor for improved gene expression (hazard ratio = 0.57, p < 0.001 for low vs. high enhancer contribution scores), with enhancer distance from the gene being the primary influencing factor, aside from circular DNA copy number (Figure 4F). These results strongly support the concept of enhancer hijacking.[12,14]

Previous studies found that ecDNA can act as a mobile super-enhancer, driving genome-wide transcriptional activation.[13] This activation depends on the structural basis of complex formation, including enhancers, transcription factors (TFs), mediators, transcription initiation proteins, and gene promoters (Figure 4G). To explore the potential mobile enhancer characteristics of HGSOC circular DNA, we performed TF binding motif enrichment analysis using HOMER,[47] based on 3,105 enhancer sequences from HGSOC circular DNAs (with 26 enhancer sequences from normal FT as background). Notably, 46.05% of enhancers on tumor circular DNA contained binding motifs for TEAD1, a key TF in the Hippo pathway that promotes cell proliferation and migration during tumor progression (Figures 4H and S6F).[48] This finding suggests that despite the high heterogeneity of circular DNA sequences, they may function as TF carriers, activating TF-specific pathways to promote tumor progression.

Transcription activation of TEs on hypomethylated ecDNA

TEs, genetic sequences capable of replicating and moving within the genome, are abundant in the human genome. In many cancers, including ovarian cancers, TEs are abnormally released from hypermethylation suppression, disrupting normal gene patterns.[49–52] Given the abundance of TE subfamilies and the advantage of long-read sequencing in resolving repeat elements, we explored the potential role of TE-amplified circular DNAs in HGSOC progression. The results showed that nearly half of the circular DNAs in HGSOC amplified TEs, including LINE, SINE, and long terminal repeats (LTRs), with absolute copy numbers reaching tens of thousands, significantly higher than those of FT samples (independent Student's t test, p < 0.0001) (Figures 5A and 5B). LINE1, the only autonomous retrotransposon in the human genome, was abnormally activated by hypomethylation.[49] As expected, we compared the methylation levels of LINE1s between HGSOC circular DNA and chromosomal genome and found that the methylation of LINE1 promoter regions on circular DNA was significantly lower (p = 0.0008; Mann-Whitney U test) (Figures 5C and S7A). Using SQuIRE to characterize TE expression from RNA-seq data,[53] we found that LINE1 expression was significantly higher in HGSOC circular DNAs than the chromosomal genome (p = 1.73e−08; Mann-Whitney U test) (Figure 5D). Similarly, non-autonomous retrotransposons, such as AluY (5mC, p = 1.67e−07; expression, p = 5.70e−26; Mann-Whitney U test) and LTRs (5mC, p = 0.019; expression, p = 4.01e−19; Mann-Whitney U test), also exhibited the same activation pattern on circular DNAs (Figures 5E and 5F). Notably, these characteristics were conserved in ecDNA and eccDNA (Figures S7B–S7D). However, these retrotransposons did not show any activation in normal FT circular DNA (Figures S7E and S7F).

To determine the impact of retrotransposon activation, we analyzed the relationship between TE-circular DNA and genomic

(G) A brief model depicting how ecDNAs function as mobile enhancers.
(H) Motif analysis based on enhancers amplified on HGSOC circular DNA (unadjusted p values from one-sided binomial test against circular DNA background sequences on normal FTs).
See also Figure S6 and Table S4.

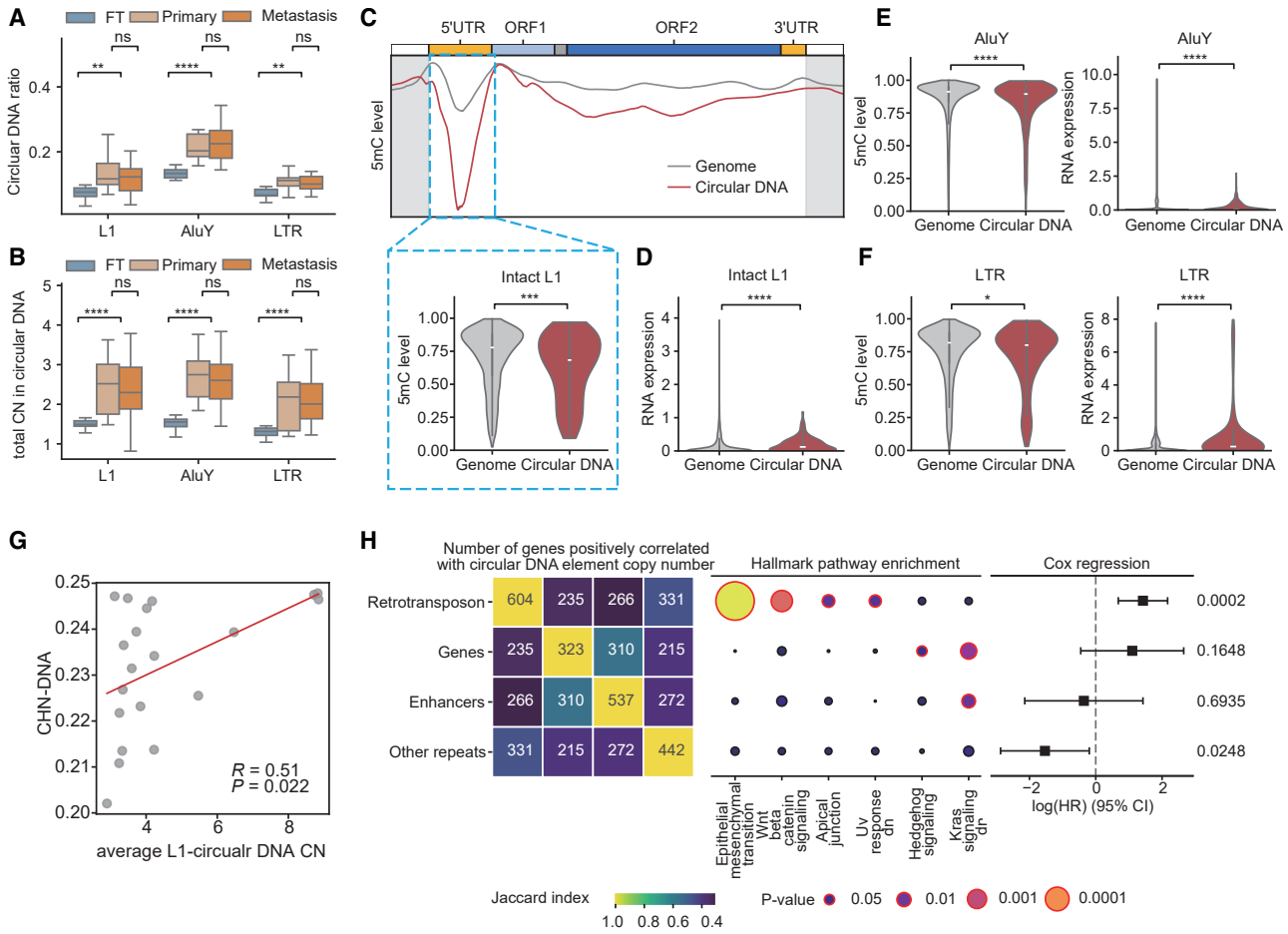


Figure 5. TEs abnormally activate transcription on hypomethylated circular DNA in HGSOC

(A) Comparison of the circular DNA ratio with TEs for FT, primary, and metastatic HGSOC samples, including L1, AluY, and long terminal repeat (LTR) (two-sided independent Student's *t* test and paired Student's *t* test; $^{**}p < 0.01$ and $^{****}p < 0.0001$).

(B) Comparison of total copy number (CN) of TEs amplified on circular DNA for FT, primary, and metastatic HGSOC samples (two-sided independent Student's *t* test and paired Student's *t* test; $^{****}p < 0.0001$).

(C and D) Methylation profile of full-length LINE1 (C). Fraction of methylation across linear and circular DNA space for HGSOC tissues. The LINE1 5′ UTR and body are highlighted in orange and blue, respectively. The bottom boxplot shows DNA methylation and RNA expression (D) of LINE1 (Mann-Whitney U test; $^{***}p < 0.001$ and $^{****}p < 0.0001$). Intact L1, length >5.9 kb.

(E and F) DNA methylation and RNA expression on TE subfamilies across linear genome and circular DNA for HGSOC tissues (Mann-Whitney U test; $^{*}p < 0.05$ and $^{****}p < 0.0001$). AluY length, >280 bp, and LTR length, >900 bp.

(G) Significant positive correlation between chromosomal CN heterogeneity score (CNH-DNA) and circular DNA-intact L1 average CN of tumor samples (Spearman's correlation, $p = 0.022$, $R = 0.51$).

(H) Left, the gene cluster positively correlated with the CN of element-amplified circular DNAs in HGSOC, including TEs, genes, enhancers, and other repeats. Middle, gene set enrichment analysis (GSEA) of the genes highlighted on the left (orange). The circle sizes represent the *p* value of the enrichment pathway. Right, multivariate Cox proportional hazards model showing the correlation between circular DNA elements and overall survival in patients with HGSOC. The error bars represent the 95% confidence intervals of the log hazard ratios. TE, transposable element.

See also Figure S7.

instability. The copy numbers of LINE1-circular DNA were significantly positively correlated with genomic instability indices, as shown by the quantized chromosomal copy-number heterogeneity score (CNH-DNA) ($R = 0.51$, $p = 0.022$; Spearman's correlation) (Figure 5G).[54,55] Based on LINE1's ability to actively insert into the genome during tumor development,[49] we profiled the somatic insertion status of LINE1 on HGSOC circular DNAs. For instance, a LINE1-circular DNA originating from chr3 showed multiple transpositions to other chromosomes, resulting in the upregulated expression of genes near the insertion loci (Figures S7G and S7H). These results suggest that transposon activation on circular DNA accelerates genomic instability and tumor heterogeneity. To better characterize the genes influenced by TE-amplified circular DNAs and their relationship to HGSOC malignancy, we identified genes positively correlated with the copy number of TE-amplified circular DNAs using gene-, enhancer-, and other repeat-circular DNAs as controls. GSEA revealed that the copy number of TE-circular DNA was

significantly positively correlated with epithelial-mesenchymal transition (EMT) pathway gene expression (Figure 5H). Multivariate analysis further showed that genes associated with TE-circular DNA copy numbers were prominently linked to shorter overall survival in patients with HGSOC (hazard ratio = 4.16, 95% confidence interval [CI] = 1.98–8.74, $p = 0.0002$) compared to genes associated with other genomic element-circular DNAs.

Taken together, the results showed that the activation of TEs amplified on hypomethylated circular DNAs contributed to genomic heterogeneity, the upregulation of EMT-related genes, and potentially increased malignancy in HGSOC.

DISCUSSION

In summary, we constructed a comprehensive profile of extrachromosomal DNA in HGSOC using long-read WGS, providing a valuable resource to resolve cancer-promoting characteristics of extrachromosomal DNA, including structures, gene elements, genetic variations, and DNA methylation. We found that compared to chromosomal amplification in HGSOC, extrachromosomal DNA suffered heavier genomic instability, which further aggravated the genetic heterogeneity of tumor cells; meanwhile, ecDNA exhibited significant DNA demethylation, especially in metastasis, which promoted cancer-associated gene expression. In addition to amplifying oncogenes, ecDNA functioned as part of a gene regulatory network, amplifying enhancers to promote gene expression through mechanisms like enhancer hijacking and mobile enhancers. Furthermore, we described that TEs were activated on hypomethylated extrachromosomal DNA, which was associated with the EMT and poor prognosis in a sample-conserved manner.

Benefiting from the genome reconstruction capabilities of long-read sequencing, we identified significantly more ecDNAs than studies based on short-read sequencing, and the PCR-free WGS library construction method allowed us to evaluate genetic variations and DNA methylation between ecDNA and chromosomal DNA. Genetic heterogeneity is a key driver of tumor evolution. Beyond the uneven distribution of ecDNA to daughter cells, the genomic instability of ecDNA provided additional driving forces for tumor evolution by promoting genomic heterogeneity. This characteristic was consistent in ecDNA and eccDNA, indicating that it is a common feature of extrachromosomal DNA. Although the PCR-free library construction method tended to detect copy-replicated ecDNA, we detected a large amount of eccDNA, which shared many common characteristics with ecDNA (Figure S8A). In the future, developing algorithms that do not rely on length screening will be more conducive to the detection of ecDNA. Besides, the content of ecDNA among patients with HGSOC displayed quantity variations. Combining this with clinical data, we found that the number of ecDNA was correlated with germline *BRCA1* mutations, which is an important DNA repair-associated gene and closely related to the malignant progression of HGSOC (Figures S8B–S8E). Whether BRCA1 mutations affect ecDNA generation remains to be further verified.

Unlike the circulatory metastasis mechanism of most tumors, HGSOC follows a typical ascites metastasis model, enabling metastatic and primary tumor cells to share many consistent features.[4] In this study, we found that metastatic and primary tumors shared many of the same ecDNAs, especially oncogene-amplified ecDNAs, such as *MDM2*, *LDHB*, *KRAS*, *DDB2*, and so on. This characteristic reflected the hereditary and selective advantages of ecDNAs. Unlike the previous study,[56] we did not find a reduction in ecDNA in HGSOC metastasis. This might be due to the malignancy of HGSOC or the sensitivity of long-read sequencing. However, compared to primary tumors, the ecDNA of metastatic tumors showed significantly lower DNA methylation, which resulted in higher expression of cancer-related genes. This characteristic represents the evolutionary trend of ecDNA in HGSOC, and whether it is conserved in other cancer types deserves further exploration.

Our research, based on the DNA hypomethylation of tumor ecDNAs, offers insights into the potential mechanisms of regulatory elements amplified on ecDNAs. By carrying enhancers alone or co-amplifying with genes, ecDNAs can drive potent gene expression to promote HGSOC progression. Using a ridge regression model, we highlighted the critical role of gene-adjacent enhancers amplified on ecDNAs in promoting gene expression. Beyond enhancer hijacking,[12,14] different circular DNA species may bind to the same TF, activating downstream pathways in an integrative manner akin to mobile enhancers.[13] Further validations, such as high-throughput chromosome conformation capture (Hi-C) and high-throughput chromatin immunoprecipitation (Hi-ChIP), are needed to confirm interactions among ecDNA, TFs, and chromosomal DNA.

Besides enhancers, retrotransposons are also affected by hypomethylation. In ovarian cancers, hypomethylated LINE1, which can restart transcription, is correlated with more aggressive histology, shorter progression-free intervals, and poorer survival.[57] Our study identified a striking pattern of significantly lower methylation of the LINE1 promoter and higher LINE1 expression on circular extrachromosomal DNAs compared to chromosomal DNA in HGSOC tissues. Additionally, our findings showed that the copy number of transposon-amplified circular DNA was highly positively correlated with EMT gene expression and shortened survival. This suggests that the activation of high-copy transposons on ecDNA may promote HGSOC tumor cell evolution. Combined with the role of transposon transposition in enhancing genomic heterogeneity, this study provides further insights into how ecDNA drives tumor evolution. Notably, this feature was conserved across clinical samples and may represent a universal oncogenic mechanism of ecDNA. As the LINE1 protein is abnormally expressed in various tumor types,[3] whether ecDNA dominates this process deserves further investigation.

Limitations of the study

Because our research is the *de novo* study of ecDNA based on a cohort's long-read WGS, the sample size is limited, and extensive public sequencing datasets cannot be integrated into our research. In the future, larger long-read sequencing datasets can be utilized to verify our findings.

Based on long-read sequencing, we discovered hundreds of ecDNAs, but only a small portion of them have been verified by DNA FISH and other methods. The accuracy of ecDNA

identified by long-read sequencing still requires systematic proof to re-examine the abundance of ecDNAs in tumors. In addition, currently, there remains a lack of effective algorithms for identifying somatic focal amplifications from long-read sequencing. Our comparison between ecDNA and chromosomal amplification may contain a few simple CNV gains, constituting a certain limitation.

This study is based on bulk sequencing and unveils a few characteristics of metastasis-related ecDNAs, except for DNA demethylation. Previous studies have shown that ecDNAs exhibit cell-type specificity and intratumoral heterogeneity.[24,58,59] Bulk sequencing may dilute cancer subclones harboring ecDNA signatures, making it challenging to analyze ecDNA's role in tumor genesis, progression, and metastasis. Future studies utilizing single-cell ecDNA sequencing, such as scEC&T-seq (a method for parallel sequencing of circular DNAs and full-length mRNA from single cells),[59] will help explore whether ecDNA promotes tumor metastasis from the perspective of lineage tracing.

RESOURCE AVAILABILITY

Lead contact

Requests for further information, resources, and reagents should be directed to and will be fulfilled by the lead contact, Shouzhen Li (lshouzhe@ustc.edu.cn).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- The sequencing data have been deposited in the Genome Sequence Archive (GSA-human: HRA008606) in the National Genomics Data Center, China National Center for Bioinformation, Beijing Institute of Genomics, Chinese Academy of Sciences.
- The custom scripts for this study have been deposited at GitHub (https://github.com/QuKunLab/ecDNA_HGSOC_TGS) and Zenodo (https://doi.org/10.5281/zenodo.16899646) and are publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2022YFA1303200 to K.Q.), the National Natural Science Foundation of China grants (T2125012 to K.Q., 32270978 to C.G., and 82473382 to Y.Z.), USTC Research Funds of the Double First-Class Initiative (YD9100002032 to K.Q. and YD9990006004 to S.L.), the Strategic Priority Research Program of Chinese Academy of Sciences (XDB0940301 to K.Q.), the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2025ZD0545600 to Y.Z.) and the Fundamental Research Funds for the Central Universities (WK9100000086 to C.G.). We thank the USTC Supercomputing Center and the School of Life Science Bioinformatics Center for providing computing resources for this project.

AUTHOR CONTRIBUTIONS

S.L. and K.Q. conceived the project. S.L., R.S., Z.L., J.F., and C.G. designed the framework. S.L. performed the wet-lab experiments (including sample collection) with help from Y.H. and Y.L. R.S. and Z.L. performed all the bioinformatics analyses with the help of C.C. and B.Z. S.L., R.S., Z.L., and K.Q. wrote the manuscript with input from all authors. K.Q. and Y.Z. supervised the project.

DECLARATION OF INTERESTS

K.Q. is the chief scientific advisor of HanGene Biotech.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT 3.5 and ChatGPT 4.0 to improve the language and readability. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human samples and ethical permission
- METHOD DETAILS
  - Nanopore whole genome library preparation and sequencing
  - RNA sequencing, whole genome sequencing (Illumina) and whole exome sequencing
  - DNA FISH
  - Circle-seq
  - Nanopore sequencing base-calling and genome alignment
  - Copy number, and structural variation calling
  - ecDNA and eccDNA detection
  - Consensus circular DNA
  - ecDNA *de novo* assembly
  - Regions of recurrent structural variation
  - Tree-shaped clustered rearrangement
  - SNV detection and mutational signature analysis
  - Whole exome sequencing data preprocess and variant calling
  - Selection of cancer-associated genes
  - ecDNA annotation plot
  - DNA methylation analysis
  - Region-set enrichment analysis using LOLA
  - RNA sequencing analysis
  - H3K27ac ChIP-seq analysis
  - Enhancer analysis
  - Transcription factor motif enrichment analysis
  - Transposable element (TE) analysis
  - Score assessing genomic instability
  - Circular DNA elements associated genes analysis
  - Overall survival analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical test

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116343.

Received: March 7, 2025
Revised: July 14, 2025
Accepted: September 5, 2025

REFERENCES

1. Siegel, R.L., Giaquinto, A.N., and Jemal, A. (2024). Cancer statistics, 2024. CA Cancer J. Clin. *74*, 12–49. https://doi.org/10.3322/caac.21820.

2. Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615. https://doi.org/10.1038/nature10166.

3. Taylor, M.S., Wu, C., Fridy, P.C., Zhang, S.J., Senussi, Y., Wolters, J.C., Cajuso, T., Cheng, W.C., Heaps, J.D., Miller, B.D., et al. (2023). Ultrasensitive Detection of Circulating LINE-1 ORF1p as a Specific Multicancer Biomarker. Cancer Discov. *13*, 2532–2547. https://doi.org/10.1158/2159-8290.CD-23-0313.

4. Wang, Y., Xie, H., Chang, X., Hu, W., Li, M., Li, Y., Liu, H., Cheng, H., Wang, S., Zhou, L., et al. (2022). Single-Cell Dissection of the Multiomic Landscape of High-Grade Serous Ovarian Cancer. Cancer Res. *82*, 3903–3916. https://doi.org/10.1158/0008-5472.CAN-21-3819.

5. Kim, H., Nguyen, N.P., Turner, K., Wu, S., Gujar, A.D., Luebeck, J., Liu, J., Deshpande, V., Rajkumar, U., Namburi, S., et al. (2020). Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. Nat. Genet. *52*, 891–897. https://doi.org/10.1038/s41588-020-0678-2.

6. Bailey, C., Pich, O., Thol, K., Watkins, T.B.K., Luebeck, J., Rowan, A., Stavrou, G., Weiser, N.E., Dameracharla, B., Bentham, R., et al. (2024). Origins and impact of extrachromosomal DNA. Nature *635*, 193–200. https://doi.org/10.1038/s41586-024-08107-3.

7. Moller, H.D., Mohiyuddin, M., Prada-Luengo, I., Sailani, M.R., Halling, J.F., Plomgaard, P., Maretty, L., Hansen, A.J., Snyder, M.P., Pilegaard, H., et al. (2018). Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. Nat. Commun. *9*, 1069. https://doi.org/10.1038/s41467-018-03369-8.

8. Mouakkad-Montoya, L., Murata, M.M., Sulovari, A., Suzuki, R., Osia, B., Malkova, A., Katsumata, M., Giuliano, A.E., Eichler, E.E., and Tanaka, H. (2021). Quantitative assessment reveals the dominance of duplicated sequences in germline-derived extrachromosomal circular DNA. Proc. Natl. Acad. Sci. USA *118*, e2102842118. https://doi.org/10.1073/pnas.2102842118.

9. Wu, S., Turner, K.M., Nguyen, N., Raviram, R., Erb, M., Santini, J., Luebeck, J., Rajkumar, U., Diao, Y., Li, B., et al. (2019). Circular ecDNA promotes accessible chromatin and high oncogene expression. Nature *575*, 699–703. https://doi.org/10.1038/s41586-019-1763-5.

10. Paulsen, T., Kumar, P., Koseoglu, M.M., and Dutta, A. (2018). Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells. Trends Genet. *34*, 270–278. https://doi.org/10.1016/j.tig.2017.12.010.

11. Liao, Z., Jiang, W., Ye, L., Li, T., Yu, X., and Liu, L. (2020). Classification of extrachromosomal circular DNA with a focus on the role of extrachromosomal DNA (ecDNA) in tumor heterogeneity and progression. Biochim. Biophys. Acta. Rev. Cancer *1874*, 188392. https://doi.org/10.1016/j.bbcan.2020.188392.

12. Morton, A.R., Dogan-Artun, N., Faber, Z.J., MacLeod, G., Bartels, C.F., Piazza, M.S., Allan, K.C., Mack, S.C., Wang, X., Gimple, R.C., et al. (2019). Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. Cell *179*, 1330–1341.e13. https://doi.org/10.1016/j.cell.2019.10.039.

13. Zhu, Y., Gujar, A.D., Wong, C.H., Tjong, H., Ngan, C.Y., Gong, L., Chen, Y.A., Kim, H., Liu, J., Li, M., et al. (2021). Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. Cancer Cell *39*, 694–707.e7. https://doi.org/10.1016/j.ccell.2021.03.006.

14. Helmsauer, K., Valieva, M.E., Ali, S., Chamorro González, R., Schöpflin, R., Röefzaad, C., Bei, Y., Dorado Garcia, H., Rodriguez-Fos, E., Puiggròs, M., et al. (2020). Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. Nat. Commun. *11*, 5823. https://doi.org/10.1038/s41467-020-19452-y.

15. Hung, K.L., Luebeck, J., Dehkordi, S.R., Colón, C.I., Li, R., Wong, I.T.L., Coruh, C., Dharanipragada, P., Lomeli, S.H., Weiser, N.E., et al. (2022). Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH. Nat. Genet. *54*, 1746–1754. https://doi.org/10.1038/s41588-022-01190-0.

16. Hung, K.L., Jones, M.G., Wong, I.T.L., Curtis, E.J., Lange, J.T., He, B.J., Luebeck, J., Schmargon, R., Scanu, E., Brückner, L., et al. (2024). Coordinated inheritance of extrachromosomal DNAs in cancer cells. Nature *635*, 201–209. https://doi.org/10.1038/s41586-024-07861-8.

17. Bergstrom, E.N., Luebeck, J., Petljak, M., Khandekar, A., Barnes, M., Zhang, T., Steele, C.D., Pillay, N., Landi, M.T., Bafna, V., et al. (2022). Mapping clustered mutations in cancer reveals APOBEC3 mutagenesis of ecDNA. Nature *602*, 510–517. https://doi.org/10.1038/s41586-022-04398-6.

18. Dharanipragada, P., Zhang, X., Liu, S., Lomeli, S.H., Hong, A., Wang, Y., Yang, Z., Lo, K.Z., Vega-Crespo, A., Ribas, A., et al. (2023). Blocking Genomic Instability Prevents Acquired Resistance to MAPK Inhibitor Therapy in Melanoma. Cancer Discov. *13*, 880–909. https://doi.org/10.1158/2159-8290.CD-22-0787.

19. Nguyen, D.D., Hooper, W.F., Liu, W., Chu, T.R., Geiger, H., Shelton, J.M., Shah, M., Goldstein, Z.R., Winterkorn, L., Helland, A., et al. (2024). The interplay of mutagenesis and ecDNA shapes urothelial cancer evolution. Nature *635*, 219–228. https://doi.org/10.1038/s41586-024-07955-3.

20. Koche, R.P., Rodriguez-Fos, E., Helmsauer, K., Burkert, M., MacArthur, I.C., Maag, J., Chamorro, R., Munoz-Perez, N., Puiggròs, M., Dorado Garcia, H., et al. (2020). Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. Nat. Genet. *52*, 29–34. https://doi.org/10.1038/s41588-019-0547-z.

21. Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods *14*, 407–410. https://doi.org/10.1038/nmeth.4184.

22. Yuen, Z.W.S., Srivastava, A., Daniel, R., McNevin, D., Jack, C., and Eyras, E. (2021). Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. Nat. Commun. *12*, 3438. https://doi.org/10.1038/s41467-021-23778-6.

23. Ng, A.W.T., McClurg, D.P., Wesley, B., Zamani, S.A., Black, E., Miremadi, A., Giger, O., Hoopen, R.T., Devonshire, G., Redmond, A.M., et al. (2024). Disentangling oncogenic amplicons in esophageal adenocarcinoma. Nat. Commun. *15*, 4074. https://doi.org/10.1038/s41467-024-47619-4.

24. Chang, L., Deng, E., Wang, J., Zhou, W., Ao, J., Liu, R., Su, D., and Fan, X. (2023). Single-cell third-generation sequencing-based multi-omics uncovers gene expression changes governed by ecDNA and structural variants in cancer cells. Clin. Transl. Med. *13*, e1351. https://doi.org/10.1002/ctm2.1351.

25. Henriksen, R.A., Jenjaroenpun, P., Sjøstrøm, I.B., Jensen, K.R., Prada-Luengo, I., Wongsurawat, T., Nookaew, I., and Regenberg, B. (2022). Circular DNA in the human germline and its association with recombination. Mol. Cell *82*, 209–217.e7. https://doi.org/10.1016/j.molcel.2021.11.027.

26. Hung, K.L., Yost, K.E., Xie, L., Shi, Q., Helmsauer, K., Luebeck, J., Schöpflin, R., Lange, J.T., Chamorro González, R., Weiser, N.E., et al. (2021). ecDNA hubs drive cooperative intermolecular oncogene expression. Nature *600*, 731–736. https://doi.org/10.1038/s41586-021-04116-8.

27. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678. https://doi.org/10.1038/s41587-020-0546-8.

28. Scheinin, I., Sie, D., Bengtsson, H., van de Wiel, M.A., Olshen, A.B., van Thuijl, H.F., van Essen, H.F., Eijk, P.P., Rustenburg, F., Meijer, G.A., et al. (2014). DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res. *24*, 2022–2032. https://doi.org/10.1101/gr.175141.114.

29. Wanchai, V., Jenjaroenpun, P., Leangapichart, T., Arrey, G., Burnham, C.M., Tümmler, M.C., Delgado-Calle, J., Regenberg, B., and Nookaew, I. (2022). CReSIL: accurate identification of extrachromosomal circular DNA from long-read sequences. Brief. Bioinform. *23*, bbac422. https://doi.org/10.1093/bib/bbac422.

30. Gao, X., Liu, K., Luo, S., Tang, M., Liu, N., Jiang, C., Fang, J., Li, S., Hou, Y., Guo, C., and Qu, K. (2024). Comparative analysis of methodologies for detecting extrachromosomal circular DNA. Nat. Commun. *15*, 9208. https://doi.org/10.1038/s41467-024-53496-8.

31. Giurgiu, M., Wittstruck, N., Rodriguez-Fos, E., Chamorro González, R., Brückner, L., Krienelke-Szymansky, A., Helmsauer, K., Hartebrodt, A., Euskirchen, P., Koche, R.P., et al. (2024). Reconstructing extrachromosomal DNA structural heterogeneity from long-read sequencing data using Decoil. Genome Res. *34*, 1355–1364. https://doi.org/10.1101/gr.279123.124.

32. Martins, F.C., Couturier, D.L., de Santiago, I., Sauer, C.M., Vias, M., Angelova, M., Sanders, D., Piskorz, A., Hall, J., Hosking, K., et al. (2022). Clonal somatic copy number altered driver events inform drug sensitivity in high-grade serous ovarian cancer. Nat. Commun. *13*, 6360. https://doi.org/10.1038/s41467-022-33870-0.

33. Smolka, M., Paulin, L.F., Grochowski, C.M., Horner, D.W., Mahmoud, M., Behera, S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., et al. (2024). Detection of mosaic and population-level structural variants with Sniffles2. Nat. Biotechnol. *42*, 1571–1580. https://doi.org/10.1038/s41587-023-02024-y.

34. Zheng, Z., Li, S., Su, J., Leung, A.W.S., Lam, T.W., and Luo, R. (2022). Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Nat. Comput. Sci. *2*, 797–803. https://doi.org/10.1038/s43588-022-00387-x.

35. Matte, I., Garde-Granger, P., Bessette, P., and Piché, A. (2019). Ascites from ovarian cancer patients stimulates MUC16 mucin expression and secretion in human peritoneal mesothelial cells through an Akt-dependent pathway. BMC Cancer *19*, 406. https://doi.org/10.1186/s12885-019-5611-7.

36. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101. https://doi.org/10.1038/s41586-020-1943-3.

37. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421. https://doi.org/10.1038/nature12477.

38. Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics *32*, 587–589. https://doi.org/10.1093/bioinformatics/btv612.

39. He, B., Zhang, C., Zhang, X., Fan, Y., Zeng, H., Liu, J., Meng, H., Bai, D., Peng, J., Zhang, Q., et al. (2021). Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. Nat. Commun. *12*, 4249. https://doi.org/10.1038/s41467-021-24425-w.

40. Pongor, L.S., Schultz, C.W., Rinaldi, L., Wangsa, D., Redon, C.E., Takahashi, N., Fialkoff, G., Desai, P., Zhang, Y., Burkett, S., et al. (2023). Extrachromosomal DNA Amplification Contributes to Small Cell Lung Cancer Heterogeneity and Is Associated with Worse Outcomes. Cancer Discov. *13*, 928–949. https://doi.org/10.1158/2159-8290.Cd-22-0796.

41. Zhong, T., zhao, H., Liu, Y., Yan, D., Li, Y., and Guo, Z. (2023). Pan-cancer characterization and clinical outcome of ecDNA-enhancer mediated transcriptional dysregulation. Preprint at bioRxiv. https://doi.org/10.1101/2023.09.06.556610.

42. Okazaki, K., Davis, D.D., and Sakano, H. (1987). T cell receptor β gene sequences in the circular DNA of thymocyte nuclei: Direct evidence for intramolecular DNA deletion in V-D-J joining. Cell *49*, 477–485. https://doi.org/10.1016/0092-8674(87)90450-8.

43. Serana, F., Chiarini, M., Zanotti, C., Sottini, A., Bertoli, D., Bosio, A., Caimi, L., and Imberti, L. (2013). Use of V(D)J recombination excision circles to identify T- and B-cell defects and to monitor the treatment in primary and acquired immunodeficiencies. J. Transl. Med. *11*, 119. https://doi.org/10.1186/1479-5876-11-119.

44. Guan, X.-Y., Sham, J.S., Tang, T.C., Fang, Y., Huo, K.K., and Yang, J.M. (2001). Isolation of a Novel Candidate Oncogene within a Frequently Amplified Region at 3q26 in Ovarian Cancer1. Cancer Res. *61*, 3806–3809.

45. Verhaak, R.G.W., Bafna, V., and Mischel, P.S. (2019). Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. Nat. Rev. Cancer *19*, 283–288. https://doi.org/10.1038/s41568-019-0128-6.

46. Corona, R.I., Seo, J.H., Lin, X., Hazelett, D.J., Reddy, J., Fonseca, M.A.S., Abassi, F., Lin, Y.G., Mhawech-Fauceglia, P.Y., Shah, S.P., et al. (2020). Non-coding somatic mutations converge on the PAX8 pathway in ovarian cancer. Nat. Commun. *11*, 2020. https://doi.org/10.1038/s41467-020-15951-0.

47. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

48. Clark, K.L., George, J.W., Przygrodzka, E., Plewes, M.R., Hua, G., Wang, C., and Davis, J.S. (2022). Hippo Signaling in the Ovary: Emerging Roles in Development, Fertility, and Disease. Endocr. Rev. *43*, 1074–1096. https://doi.org/10.1210/endrev/bnac013.

49. Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., 3rd, Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K., et al. (2012). Landscape of Somatic Retrotransposition in Human Cancers. Science *337*, 967–971. https://doi.org/10.1126/science.1222077.

50. Nguyen, T.H.M., Carreira, P.E., Sanchez-Luque, F.J., Schauer, S.N., Fagg, A.C., Richardson, S.R., Davies, C.M., Jesuadian, J.S., Kempen, M.J.H.C., Troskie, R.L., et al. (2018). L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution. Cell Rep. *23*, 3730–3740. https://doi.org/10.1016/j.celrep.2018.05.090.

51. Payer, L.M., and Burns, K.H. (2019). Transposable elements in human genetic disease. Nat. Rev. Genet. *20*, 760–772. https://doi.org/10.1038/s41576-019-0165-8.

52. Tang, Z., Steranka, J.P., Ma, S., Grivainis, M., Rodić, N., Huang, C.R.L., Shih, I.M., Wang, T.L., Boeke, J.D., Fenyö, D., and Burns, K.H. (2017). Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. Proc. Natl. Acad. Sci. USA *114*, E733–E740. https://doi.org/10.1073/pnas.1619797114.

53. Yang, W.R., Ardeljan, D., Pacyna, C.N., Payer, L.M., and Burns, K.H. (2019). SQuIRE reveals locus-specific regulation of interspersed repeat expression. Nucleic Acids Res. *47*, e27. https://doi.org/10.1093/nar/gky1301.

54. Liu, H., Gao, J., Feng, M., Cheng, J., Tang, Y., Cao, Q., Zhao, Z., Meng, Z., Zhang, J., Zhang, G., et al. (2024). Integrative molecular and spatial analysis reveals evolutionary dynamics and tumor-immune interplay of in situ and invasive acral melanoma. Cancer Cell *42*, 1067–1085.e1011. https://doi.org/10.1016/j.ccell.2024.04.012.

55. van Dijk, E., van den Bosch, T., Lenos, K.J., El Makrini, K., Nijman, L.E., van Essen, H.F.B., Lansu, N., Boekhout, M., Hageman, J.H., Fitzgerald, R.C., et al. (2021). Chromosomal copy number heterogeneity predicts survival rates across cancers. Nat. Commun. *12*, 3188. https://doi.org/10.1038/s41467-021-23384-6.

56. Kim, H., Kim, S., Wade, T., Yeo, E., Lipsa, A., Golebiewska, A., Johnson, K.C., An, S., Ko, J., Nam, Y., et al. (2024). Mapping extrachromosomal DNA amplifications during cancer progression. Nat. Genet. *56*, 2447–2454. https://doi.org/10.1038/s41588-024-01949-7.

57. Pattamadilok, J., Huapai, N., Rattanatanyong, P., Vasurattana, A., Triratanachat, S., Tresukosol, D., and Mutirangura, A. (2008). LINE-1 hypomethylation level as a potential prognostic factor for epithelial ovarian cancer. Int. J. Gynecol. Cancer *18*, 711–717. https://doi.org/10.1111/j.1525-1438.2007.01117.x.

58. Chen, J.P., Diekmann, C., Wu, H., Chen, C., Della Chiara, G., Berrino, E., Georgiadis, K.L., Bouwman, B.A.M., Virdi, M., Harbers, L., et al. (2024). scCircle-seq unveils the diversity and complexity of extrachromosomal

circular DNAs in single cells. Nat. Commun. *15*, 1768. https://doi.org/10.1038/s41467-024-45972-y.

59. Chamorro González, R., Conrad, T., Stöber, M.C., Xu, R., Giurgiu, M., Rodriguez-Fos, E., Kasack, K., Brückner, L., van Leen, E., Helmsauer, K., et al. (2023). Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells. Nat. Genet. *55*, 880–890. https://doi.org/10.1038/s41588-023-01386-y.
60. Wick, R.R., Judd, L.M., and Holt, K.E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. *20*, 129. https://doi.org/10.1186/s13059-019-1727-y.
61. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.
62. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 31. https://doi.org/10.1186/s13059-016-0893-4.
63. Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics *34*, 867–868. https://doi.org/10.1093/bioinformatics/btx699.
64. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.
65. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. Bioinformatics *30*, 2811–2812. https://doi.org/10.1093/bioinformatics/btu393.
66. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. *37*, 540–546. https://doi.org/10.1038/s41587-019-0072-8.
67. Ewing, A.D., Smits, N., Sanchez-Luque, F.J., Faivre, J., Brennan, P.M., Richardson, S.R., Cheetham, S.W., and Faulkner, G.J. (2020). Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. Mol. Cell *80*, 915–928.e5. https://doi.org/10.1016/j.molcel.2020.10.024.
68. Wang, Y., Wang, M., Djekidel, M.N., Chen, H., Liu, D., Alt, F.W., and Zhang, Y. (2021). eccDNAs are apoptotic products with high innate immunostimulatory activity. Nature *599*, 308–314. https://doi.org/10.1038/s41586-021-04009-w.
69. Imielinski, M., Guo, G., and Meyerson, M. (2017). Insertions and Deletions Target Lineage-Defining Genes in Human Cancers. Cell *168*, 460–472. e14. https://doi.org/10.1016/j.cell.2016.12.025.
70. Liu, Y., Sun, J., and Zhao, M. (2017). ONGene: A literature-based database for human oncogenes. J. Genet. Genomics *44*, 119–121. https://doi.org/10.1016/j.jgg.2016.12.004.
71. Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D., and Ferretti, V. (2019). The International Cancer Genome Consortium Data Portal. Nat. Biotechnol. *37*, 367–369. https://doi.org/10.1038/s41587-019-0055-9.
72. Ewing, A., Meynert, A., Silk, R., Aitken, S., Bendixsen, D.P., Churchman, M., Brown, S.L., Hamdan, A., Mattocks, J., Grimes, G.R., et al. (2024). Divergent trajectories to structural diversity impact patient survival in high grade serous ovarian cancer. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.12.575376.
73. Masoodi, T., Siraj, S., Siraj, A.K., Azam, S., Qadri, Z., Parvathareddy, S.K., Tulbah, A., Al-Dayel, F., AlHusaini, H., AlOmar, O., et al. (2020). Genetic heterogeneity and evolutionary history of high-grade ovarian carcinoma and matched distant metastases. Br. J. Cancer *122*, 1219–1230. https://doi.org/10.1038/s41416-020-0763-4.
74. Kotnik, E.N., Mullen, M.M., Spies, N.C., Li, T., Inkman, M., Zhang, J., Martins-Rodrigues, F., Hagemann, I.S., McCourt, C.K., Thaker, P.H., et al. (2023). Genetic characterization of primary and metastatic high-grade serous ovarian cancer tumors reveals distinct features associated with survival. Commun. Biol. *6*, 688. https://doi.org/10.1038/s42003-023-05026-3.
75. Vias, M., Morrill Gavarró, L., Sauer, C.M., Sanders, D.A., Piskorz, A.M., Couturier, D.L., Ballereau, S., Hernando, B., Schneider, M.P., Hall, J., et al. (2023). High-grade serous ovarian carcinoma organoids as models of chromosomal instability. eLife *12*, e83867. https://doi.org/10.7554/eLife.83867.
76. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523. https://doi.org/10.1038/s41467-019-09234-6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
High-grade serous ovarian cancer tissues, fully anonymized	The First Affiliated Hospital of University of Science and Technology of China	N/A
Fallopian tube tissues, fully anonymized	The First Affiliated Hospital of University of Science and Technology of China	N/A
Peripheral blood mononuclear cells, fully anonymized	The First Affiliated Hospital of University of Science and Technology of China	N/A
Critical commercial assays		
Oxford Nanopore Ligation Sequencing Kit	Oxford Nanopore Technologies	SQK-Q20EA; SQK-LSK114
R10.4.1 flow cells	Oxford Nanopore Technologies	FLO-PRO112
MagAttract HMW DNA kit	QIAGEN	67563
g-TUBE	Covaris	N/A
RNAlater solution	Invitrogen	AM7024
TRIzol Reagent	Invitrogen	15596026CN
Ficoll-Paque Plus	Solarbio	P8610
DNeasy Blood & Tissue Kit	QIAGEN	69504
QuarHyb Super Reagent Kit Pro	Dynegene	NC1006A
QuarXeq Human All Exon Probes 3.0	Dynegene	NY1009
QIAGEN Plasmid Plus Midi Kit	QIAGEN	12943
TruePrep DNA Library Prep Kit V2	Vazyme	TD501
MYC gene amplification probe	SpatialFISH	DFPH2003
MDM2 gene amplification probe	SpatialFISH	DFPH2018
Long-read sequencing	Novogene	N/A
Short-read sequencing	Sequanta Technologies	N/A
Mss1	Thermo Scientific	ER1341
ATP-dependent Plasmid Safe DNase	Lucigen	E3101K
SPRIselect beads	Beckman	B23318
phi29 DNA polymerase	New England Biolabs	M0269L
inorganic pyrophosphatase	New England Biolabs	M2403S
Exo-Resistant Random Primer	Thermo Scientific	SO181
25mM dNTPs	Thermo Scientific	R1121
Deposited data		
Raw and analyzed data	This paper	GSA-Human: HRA008606
H3K27ac ChIP-seq data for HGSOC tissues	Corona et al.[46]	GEO: GSE121103
Software and algorithms		
CReSIL	Wanchai et al.[29]	https://github.com/visanuwan/cresil
Decoil	Giurgiu et al.[31]	https://github.com/madagiurgiu25/decoil-pre
QDNAseq (v1.30.0)	Scheinin et al.[28]	https://github.com/ccagc/QDNAseq
Sniffles2 (v2.2)	Smolka et al.[33]	https://github.com/fritzsedlazeck/Sniffles
Clair3 (v1.0.4)	Zheng et al.[34]	https://github.com/HKU-BAL/Clair3
Guppy	Wick et al.[60]	https://nanoporetech.com/community
ANNOVAR	Wang et al.[61]	https://github.com/WGLab/doc-ANNOVAR
deconstructSigs (v1.9.0)	Rosenthal et al.[62]	https://github.com/raerose01/deconstructSigs

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mosdepth	Pedersen et al.[63]	https://github.com/brentp/mosdepth
deepTools (v3.5.1)	Ramírez et al.[64]	https://github.com/deeptools/deepTools
Circlize (v0.4.15)	Gu et al.[65]	https://github.com/jokergoo/circlize
Flye v2.9.2-b1786	Kolmogorov et al.[66]	https://github.com/mikolmogorov/Flye
HOMER (v4.11)	Heinz et al.[47]	http://homer.ucsd.edu/homer/
SQuIRE (v0.9.9.9a)	Yang et al.[53]	https://github.com/wyang17/SQuIRE
TLDR	Ewing et al.[67]	https://github.com/adamewing/tldr
Custom scripts for this study	This paper	https://github.com/QuKunLab/ecDNA_HGSOC_TGS; https://doi.org/10.5281/zenodo.16899646

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human samples and ethical permission

This study was approved by the medical research ethics committee of The First Affiliated Hospital of University of Science and Technology of China (approval No. 281 2022KY), and all patients provided signed informed consent accordingly before the surgery.

Patients with high-grade serous ovarian cancer (HGSOC) were enrolled only if they had not received tumor-related treatments before surgery. Peripheral blood samples (<5 mL) were collected during preoperative examination using EDTA anticoagulant tubes, and primary and paired metastatic tumor tissues were collected simultaneously during the surgery. Fallopian tubes were obtained from patients who underwent bilateral salpingo-oophorectomy for benign gynecologic diseases and risk-reducing salpingo-oophorectomy surgery. Normal fallopian tube tissues were further conformed by CNV analysis.

METHOD DETAILS

Nanopore whole genome library preparation and sequencing

Genomic DNA were extracted from the tissues frozen in liquid nitrogen using MagAttract HMW DNA kit (Qiagen 67563). Then the Covaris g-TUBE was administrated to shear genomic DNA into selected 20kb fragment sizes, followed by quality control with pulsed field electrophoresis. Libraries for sample OC1-3 were constructed using SQK-Q20EA kit (Oxford Nanopore Technologies) according to the manufacturer's instruction, and SQK-LSK114 was adopted for sample OC4-12 and N1-6. Sequencing was performed on R10.4.1 flow cells (FLO-PRO112) on the PromethION platform by Novogene service provider.

RNA sequencing, whole genome sequencing (Illumina) and whole exome sequencing

For RNA sequencing, tissues were shortly stored in RNAlater solution (Invitrogen), and after crushed, total RNA was extracted by TRIzol reagent (Invitrogen) followed by library preparation and sequencing on the Novaseq 6000 PE150 platform (Illumina) by Sequanta Technologies.

For next-generation whole genome sequencing, peripheral blood mononuclear cells (PBMCs) were prepared by Ficoll-Paque density gradient centrifugation (Solarbio, P8610), and genomic DNA was extracted by DNeasy Blood & Tissue Kit (Qiagen, 69504) according to the manufacturer's instruction. Library preparation and sequencing were also conducted by Sequanta Technologies on the Novaseq 6000 PE150 platform.

For whole exome sequencing, frozen tissues were crushed into powder and genomic DNA was extracted by DNeasy Blood & Tissue Kit (Qiagen, 69504) according to the manufacturer's instruction. Libraries were constructed using QuarHyb Super Reagent Kit Pro and QuarXeq Human All Exon Probes 3.0 (Dynegene, Shanghai) by Bioten Biotechnology, and sequenced on the BGI DNBSEQ-T7 platform.

DNA FISH

Fresh tissues frozen in OCT were sectioned and fixed with 4% paraformaldehyde at room temperature for 10 min. The samples were washed with PBST followed by denaturation with 100% methanol. And the pepsin was used to permeate cells and expose DNA. Then The samples were dehydrated in ascending ethanol series (70%, 85%, 95%, 100%) for 1 min. After adding probe hybridization buffer, samples and probes were co-denatured at 82°C for 10 min, and then incubated at 37°C for 16 h. Finally, after washing with 0.4×SSC, DAPI staining and sealing were performed. Images were captured by Leica stellaris5 confocal microscope. For *MYC* and *MDM2* FISH, DFPH2003 and DFPH2018 kits were adopted, respectively; while the probes for *DDB2* FISH were customized.

Circle-seq

Circle-seq referred to the published protocol with some modifications.[68] Briefly, the tissue was cut into pieces and then crude circular DNA was extracted by QIAGEN Plasmid Plus Midi Kit according to the manufacturer's instructions. Mitochondrial DNA and linear

DNA were digested with Mss1 and ATP-dependent Plasmid Safe DNase at 37°C for 48 h, and 1.5× SPRIselect beads were used to purify DNA. Subsequently, Rolling Cycle Amplification was carried out using phi29 DNA polymerase for 16 h, and 0.4× SPRIselect beads were used to purify DNA. Illumina sequencing libraries for circular DNA were prepared by Tn5 transposon-based tagmentation with TruePrep DNA Library Prep Kit V2 for Illumina according to the manufacturer's instructions. Finally, the barcoded libraries were pooled together and sequenced with the Illumina NovaSeq 6000 platform in 150-bp paired-end mode (Sequanta Technologies). Sequencing read coverage per 50-bp bin was calculated using deeptools 'bamCoverage' (v3.5.1) with default values.[64]

Nanopore sequencing base-calling and genome alignment

Base-calling and DNA methylation detection from raw ONT sequencing data was performed using Guppy (v6.5.7) with high-accuracy models for R10.4 or R10.4.1.[60] Quality control of base-called raw reads was performed using pycoQC (v2.5.2). Reads with mean quality <10 was excluded. Reads were then aligned to hg38 human reference genome using Guppy integrative minimap2 (https://github.com/lh3/minimap2) with default parameters. After alignment, BAM files were sorted and indexed for downstream analyses, including copy number variation (CNV), ecDNA, single-nucleotide variant (SNV), and structural variation (SV) detection.

Copy number, and structural variation calling

CNVs were calculated using QDNAseq (v1.30.0) with bin sizes of 15 kb and 100 kb.[28] Two metastatic tissues (OC1M, OC5M) were excluded from subsequent analysis due to sparse tumor components under CNV analysis, which did not affect the statistical conclusions. Finally, 6 normal FT and 22 tumor samples were selected for the following analysis.

Structural Variant detection with ONT mapped libraries was performed on each sample using Sniffles2 (v2.2) with the following optional parameters: General SV (all Modes)+ Mosaic SV (low-frequency, 2–30% variant allele frequency) Calling.[33] Mosaic SVs were identified with parameters "–mosaic-qc-coverage-max-change-frac 0.8 –mosaic-af-min 0.02 –mosaic-af-max 0.3".

ecDNA and eccDNA detection

Aligned long-read WGS reads were used as input data for CReSIL to identify potential circular DNA regions with the following command 'cresil identify_wgls',[29] and for Decoil 1.1.3 to identify ecDNA regions.[31] Mosdepth was used to calculate circular DNA length-weighted copy number.[63] eccDNA was defined as circular DNA identified by CReSIL with length <100 kb and length-weighted copy number <3. ecDNA was defined as circular DNA identified by CReSIL or Decoil with length >100 kb and length-weighted copy number >3. Two ecDNAs were merge if one of them has more than 50% region overlapped with the other. Complex circular DNAs were defined as those assembled from multiple genomic fragments, while single-locus circular DNAs were classified as simple ecDNAs. Final circular DNA lists for analysis were filtered by removing all breakpoints overlapping with ENCODE blacklisted regions (https://sites.google.com/site/anshulkundaje/projects/blacklists).

Consensus circular DNA

Circular DNA with 80% sequence overlapped between primary and metastatic tissues is considered conserved circular DNA. In paired samples, if one contains 95% sequence (at least 40% overlap with each other) of the other, it is also considered conserved circular DNA. Calculate the proportion of conserved circular DNA in each sample to all circular DNA.

ecDNA *de novo* assembly

To carry out ecDNA *de-novo* assembly, we extracted reads mapping to ecDNA regions with samtools view and Seqtk subseq v1.4-r122 and used Flye v2.9.2-b1786 to assemble the circular contig.[66] We transferred genomic coordinates and genomic features of hg38 reference genome to the circular contig using minimap2 and liftOver (v362, https://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/).

Regions of recurrent structural variation

Following previous procedure,[19] we used a Gamma-Poisson model, as implemented in the R package FishHook,[69] to discover regions of recurrent structural variation. The genome was partitioned into 100kb non-overlapping bins, and the union of mosaic SV breakpoints from each tumor sample was used as input. Covariates were added to model the background mutation rate, including: nucleotide frequency, dinucleotide frequency, trinucleotide frequency, H3K4me3 marks [ENCODE accession: ENCFF112KXI], H3K27ac marks [ENCFF554TER], H3K4me1 marks [ENCFF030RVT]; DNase hypersensitivity sites [ENCFF590ZTY]; replication timing (https://github.com/skandlab/MutSpot/tree/master/features/Ch38), fragile sites [HGNC 2021], and RepeatMasker, LINE, SINE, LTR, simple repeat, and DNA transposon annotations from UCSC. An FDR-adjusted (Benjamani-Hochberg) p-value cutoff of 0.25 was used to nominate significant breakpoint hits. Oncogenes were derived from a combination of the ONGene database.[70]

Tree-shaped clustered rearrangement

Following previous study,[20] mosaic SVs were used as input data to find regions of clustered, tree-shaped rearrangement pattern which having three or more interchromosomal rearrangements within a 4-Mb sliding window. All chromosomes with >25 interchromosomal rearrangements were removed.

SNV detection and mutational signature analysis

SNVs were detected using Clair3 (v 1.0.4) with the '—min_mq 20' option from BAM files of tumor tissues and matched PBMC data from the same patient as a reference to filter germline SNVs.[34] The patients without matched PBMC data (OC11 OC12) were excluded for SNV detection. SNVs were excluded with Depth<= 15, QUAL $\leq$ 5, and reads support <3. Somatic SNVs were then annotated for gene-coding status using ANNOVAR.[61] Genome regions were defined as whole genome regions except for chromosome Y. Amplified regions were defined as CNV regions with copy number >2.5 and not overlapped with ecDNA/eccDNA regions.

Mutational signatures were analyzed using a non-negative matrix factorization–based tool deconstructSigs v1.9.0 with parameters ''contexts.needed = TRUE, signature.cutoff = 0.005, tri.counts.method = 'default'',[62] with single base substitution (SBS) signatures V3.4 in the 96 trinucleotide contexts as reference. To characterize the broad spectrum of SBS signatures, we calculated the TMB proportions of SBS signatures within ecDNA regions and within whole genome region and chromosomal amplified regions.

Whole exome sequencing data preprocess and variant calling

WES reads were trimmed and filtered by fastp v0.24.0(-l 30 -f 9 –adapter_fasta adapter.fa). Reads pass filter were mapped with BWA MEM v0.7.19-r1273 to the hg38 reference genome. Small variants from WES were identified with GATK HaplotypeCaller and GenotypeGVCFs v4.6.2.0 and the common mutations between WES and Nanopore sequencing were identified by bcftools isec v1.21.

Selection of cancer-associated genes

We used a list of cancer-associated genes to annotate CNV and SNV, including oncogenes, tumor suppressor genes (TSGs), and some cancer-driver genes, was compiled from ICGC Data Portal databases,[71] and previous studies.[72–75] Oncogenes were derived from a combination of the ONGene database to annotate ecDNA.[70] Only intact gene on ecDNA was considered. Immunomodulatory genes mapped to circular DNA selected from gene set had an immunomodulatory function (GO terms: 0006968, 0002228, 0042267, 0001906, 0001909, 0002698, 0001910, 0031341, 0002695, 0050866, 0051250, 0050777).

ecDNA annotation plot

The ecDNA genes were annotated when they overlap with the hg38 annotation gtf file for whole genebody. The circles diagram was plotted using R package Circlize (v0.4.15).[65] The circle annotations from outside to inside are genomic location, transposable element (LINE1 (blue), AluY (gray), LTR (green)), enhancer, DNA methylation, gene expression, and gene.

DNA methylation analysis

The 5-methylcytosine and 5-hydroxymethylcytosine status of whole genome cytosine sites were called from Guppy aligned BAM files using Modkit pileup (v0.4.4). Methylation level for a region r was calculated as:

$$Methylation\ level_r = \frac{\sum_{i \in r} N_{mod,i}}{\sum_{i \in r} N_{valid,i}}$$

N_{mod} is number of modified calls passing filters and N_{valid} is the valid coverage as defined in Modkit. Regions with valid coverage lower than 0.3 * region length were filtered. Relative methylation level for a peak p was calculated as:

$$Relative\ methylation\ level_p = methylation\ level_p - methylation\ level_{p\_flank}$$

The flank region p_flank for a peak p was defined as upstream and downstream 10 kb of the peak region.

The comparison of methylation level on genome, amplified and circular DNA regions were performed by comparing relative methylation level of peaks from H3K27ac ChIP-seq data in particular regions.

Region-set enrichment analysis using LOLA

To identify shared genomic patterns among circular DNA regions, the LOLA software was used to perform region-set enrichment analysis against circular DNA region sets.[38] The ecDNA regions on normal fallopian tubes or on HGSOC tissues were used as the LOLA query set, and the set of all circular DNA regions were used as background ('universe'). UCSC features and encode segmentations were obtained from the LOLA Core database used in the region set enrichment analysis, and enrichments with a $p < 0.05$ were considered significant.

RNA sequencing analysis

FastQC (v0.11.9) was conducted to assess the RNA-Seq data quality. Low quality reads were trimmed using Trimmomatic (v0.39) with parameters: ILLUMINACLIP: TruSeq3-PE.fa:2:30:10:2:True HEADCROP:10 LEADING:3 TRAILING:3 MINLEN:36. Processed read pairs were aligned to the human reference genome (GRCh38) using STAR (v2.7.10a). Mapped read counts per gene were measured using featureCounts (v2.0.1) and GENCODE version 41 gene annotation. TE expression (FPKM) was quantified by SQuIRE (v0.9.9.9a) (https://github.com/wyang17/SQuIRE) following the default pipeline.[53]

H3K27ac ChIP-seq analysis

H3K27ac ChIP-seq data for HGSOC tissues were previously published under Gene Expression Omnibus accession GSE121103.[46] ChIP-seq data were trimmed by trimmomatic (v0.39) with parameters: ILLUMINACLIP: TruSeq3-SE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. The clean reads were aligned to the hg38 genome assembly using Bowtie2 (v2.5.1), and removed duplicates, unmapped reads using sambamba (v0.6.6). Peaks were detected with MACS2 (v2.2.7.1), using a matched input DNA sample as control with a threshold for significant enrichment of $p < 1 \times 10^{-9}$. Peaks lists were filtered to remove all peaks overlapping ENCODE blacklisted regions. The consensus peak regions for HGSOC were present in at least three (out of five) HGSOC samples. ChIP-seq signal was converted to the bigwig format for visualization using deepTools bamCoverage (v3.5.1) with the following parameters: –binSize 10 –normalizeUsing RPGC –effective GenomeSize 2913022398 –extendReads 75.

Enhancer analysis

Enhancer annotations were obtained using HOMER (v4.11).[47] Enhancers on circular DNA were categorized based on their association with genes: enhancer+/gene+ (enhancer and gene co-amplified) or enhancer+/gene– (enhancer only). Pathway enrichment analysis was performed using MetaScape with the list of genes on circular DNA or genes associated with enhancers on circular DNA as input.[76] Circular DNA was annotated by several enhancer features. These features included the number of enhancers (enhancer number), the log-transformed distance to the nearest enhancer (distance from gene), the copy number of circular DNA (circular DNA copy number), and the DNA methylation level of the closest enhancer (DNA methylation). The enhancer metrics were then averaged for each gene to generate a comprehensive enhancer profile. We calculated enhancer contribution score for each gene of each enhancer+/gene+ circular DNA by using an optimized Ridge regression model. Hazard ratios (HR) were computed using a Cox proportional hazards regression model from Survminer (v0.5.0) R package (gene expression data available for enhancer +/gene+ circular DNA).

Transcription factor motif enrichment analysis

To explore the potential mobile characteristic on HGSOC circular DNA, we extracted all enhancer sequences on circular DNA as fasta and applied HOMER (v4.11) findMotifs.pl to search specifically enriched motifs.[47] Enhancer sequences within HGSOC circular DNA were used as the target sequences against within normal FT circular DNA as normalized background sequences.

Transposable element (TE) analysis

Reference TE locations were derived from the RepeatMasker (http://www.repeatmasker.org/).out files available for hg38 from the UCSC Genome Browser. LINE1 elements longer than 1.2kb (full-length LINE1>5.9kb), AluY elements longer than 280 bp, and LTR elements longer than 900 bp were considered. TLDR was used to detect TE insertions from ONT BAM files using full-length LINE1 sequence on circular DNA as reference elements(-e/–elts).[67]

Score assessing genomic instability

Following previously studies,[54,55] CNH-DNA score of each tumor sample is calculated as genomic instability indice. CNH-DNA score is a scalable metric designed to quantify intertumoral heterogeneity and genomic instability based on copy-number profiling.

Circular DNA elements associated genes analysis

To explore the associated genes of different elements on circular DNA, we applied partial Pearson correlation test between circular DNA elements (TEs, genes, enhancers and other repeats) average copy number and genes expression ($\log_2$(TPM+1)) using sequencing depth as covariables. Genes with p-value <0.01 and R > 0.5 were classified as circular DNA elements associated genes and further performed HALLMARK pathway enrichment using GSEApy (v1.1.4).

Overall survival analysis

The clinical survival data and paired mRNA expression array data was downloaded from KM plotter. To investigate the impact of ecDNA elements associated genes on overall survival, we performed a Cox proportional hazards regression model to determine the hazards ratios for different elements associated genes using Lifelines python package (v0.30.0).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical test

All statistical tests used to analyze the data are specifically described in the main text and figure legends or STAR Methods section. Statistical tests included two-sided *t* test, two-sided Mann-Whitney U test, two-sided Wilcoxon rank-sum test, two-sided Fisher's exact test, and Pearson or Spearman cor.test. Unless otherwise specified, boxes show the median and the interquartile range (IQR); the lower whisker indicates Q1 – 1.5 × the IQR; the upper whisker indicates Q3 + 1.5 × the IQR. Significant p values were indicated as follows: $p \leq 0.05$ (*), $p \leq 0.01$(**) and $p \leq 0.001$(***), $p \leq 0.0001$(****).

Cell Reports, Volume *44*

Supplemental information

Landscape of extrachromosomal DNA characteristics in high-grade serous ovarian cancer via long-read sequencing

Ruoming Sun, Zongkai Li, Yi Liu, Yanbing Hou, Bowen Zhao, Chunpeng Chen, Jingwen Fang, Chuang Guo, Ying Zhou, Kun Qu, and Shouzhen Li

Supplemental materials

The file includes:

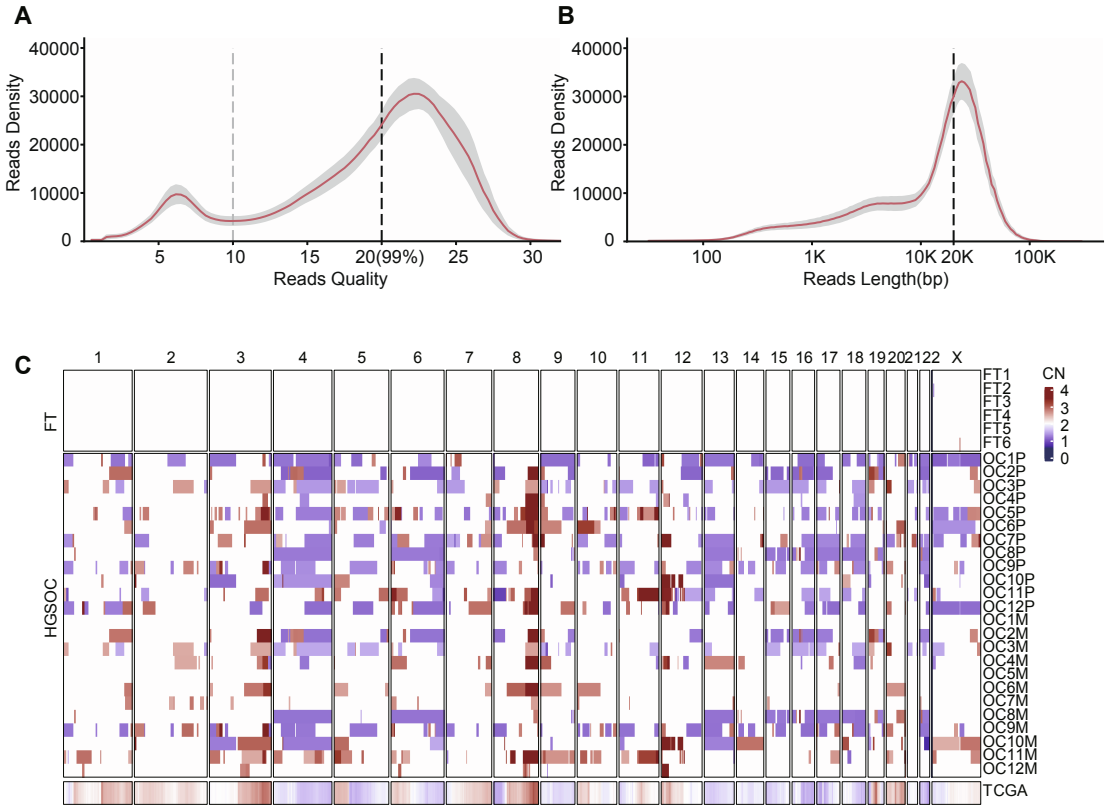


Figure S1. Summary of the long-read sequencing data.

(A) The sequencing reads quality of long-read data obtained from 30 samples via nanopore sequencer. (B) The sequencing reads length of long-read data obtained from 30 samples via nanopore sequencer. The vertical dashed line indicates the average reads quality and length across all samples in the cohort. (C) Global CNV patterns inferred from long-read WGS data of FT and HGSOC tissues. Each row of the heatmap represents a sample. Each column represents a chromosome. The barplot on the top represents the CNV frequency of all samples within 100kb bin. The heatmap on the bottom represents the CNV frequency of TCGA OV within 10kb bin. Red represents CN gain and blue represents CN loss.

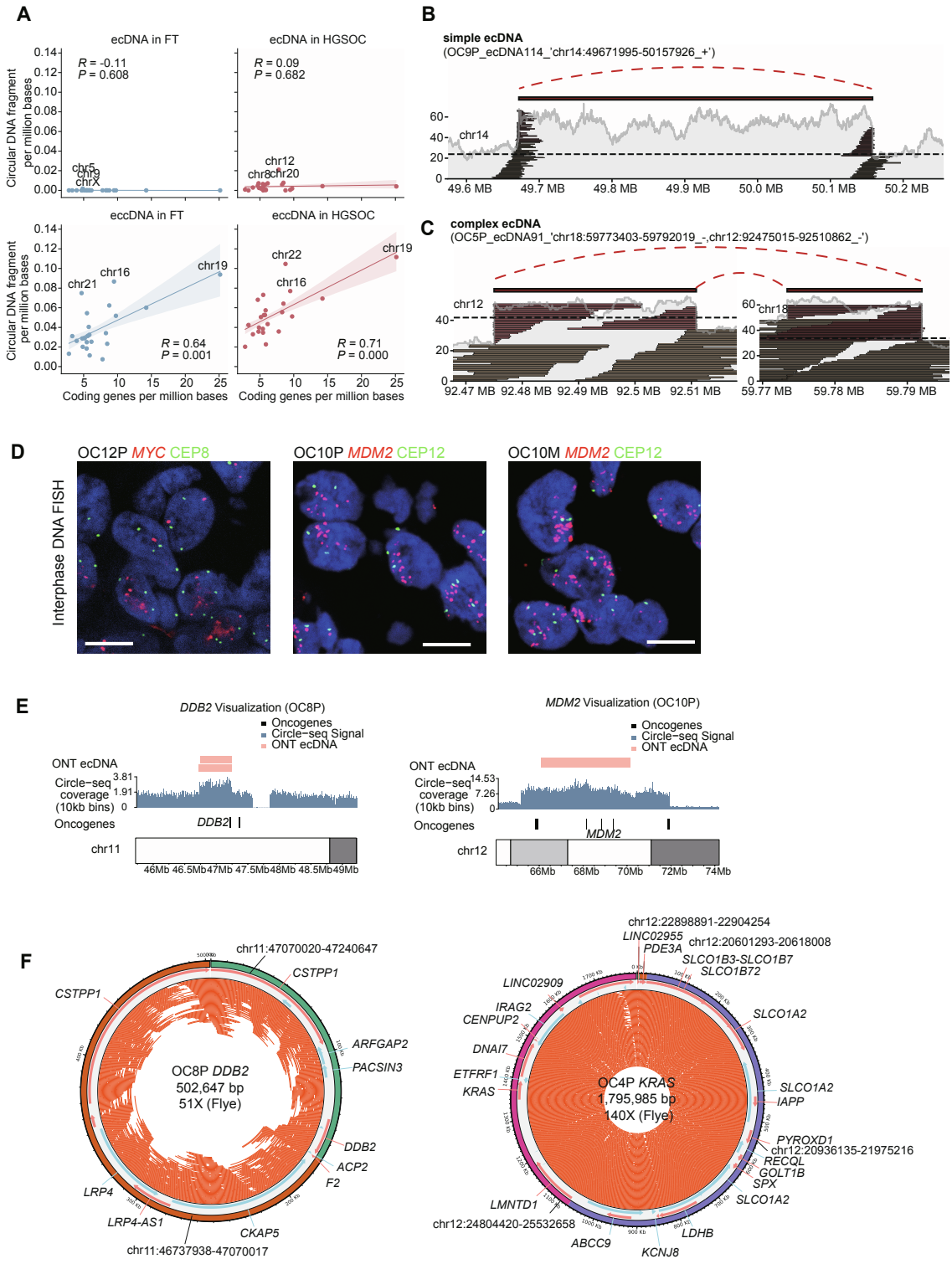


Figure S2. Validation of the presence of ecDNA.

(A) Significant positive correlation between relative number of eccDNA identified per chromosome (Y chromosome and mitochondria not included) and number of coding genes (HGSOC eccDNA, $P < 0.0001$, $R = 0.71$; FT eccDNA, $P = 0.001$, $R = 0.64$) compared with no positive correlation for ecDNA (HGSOC ecDNA, $P = 0.682$, $R = 0.09$; FT ecDNA, $P = 0.608$, $R = -0.11$).

(B) Distinct rearrangement patterns of ecDNA. The panel demonstrates a simple ecDNA pattern, where the two break sites of a single focal amplification are assembled.
(C) The panel demonstrates a complex ecDNA pattern, with multiple fragments connected at break sites marked with red arches. Expected diploid levels of sequencing depths are marked by horizontal dashed lines. The position of split reads is marked with red lines.
(D) Representative interphase DNA FISH of matched frozen HGSOC tissues showing oncogene-amplified ecDNA against centromere labeling. Scale bars, 10 μm.
(E) Validation of the presence of *DDB2* and *MDM2* on ecDNA by Circle-Seq coverage for matched frozen HGSOC tissues.
(F) Cyclic assembly graphs of OC8P and OC4P showing position of genes amplified.

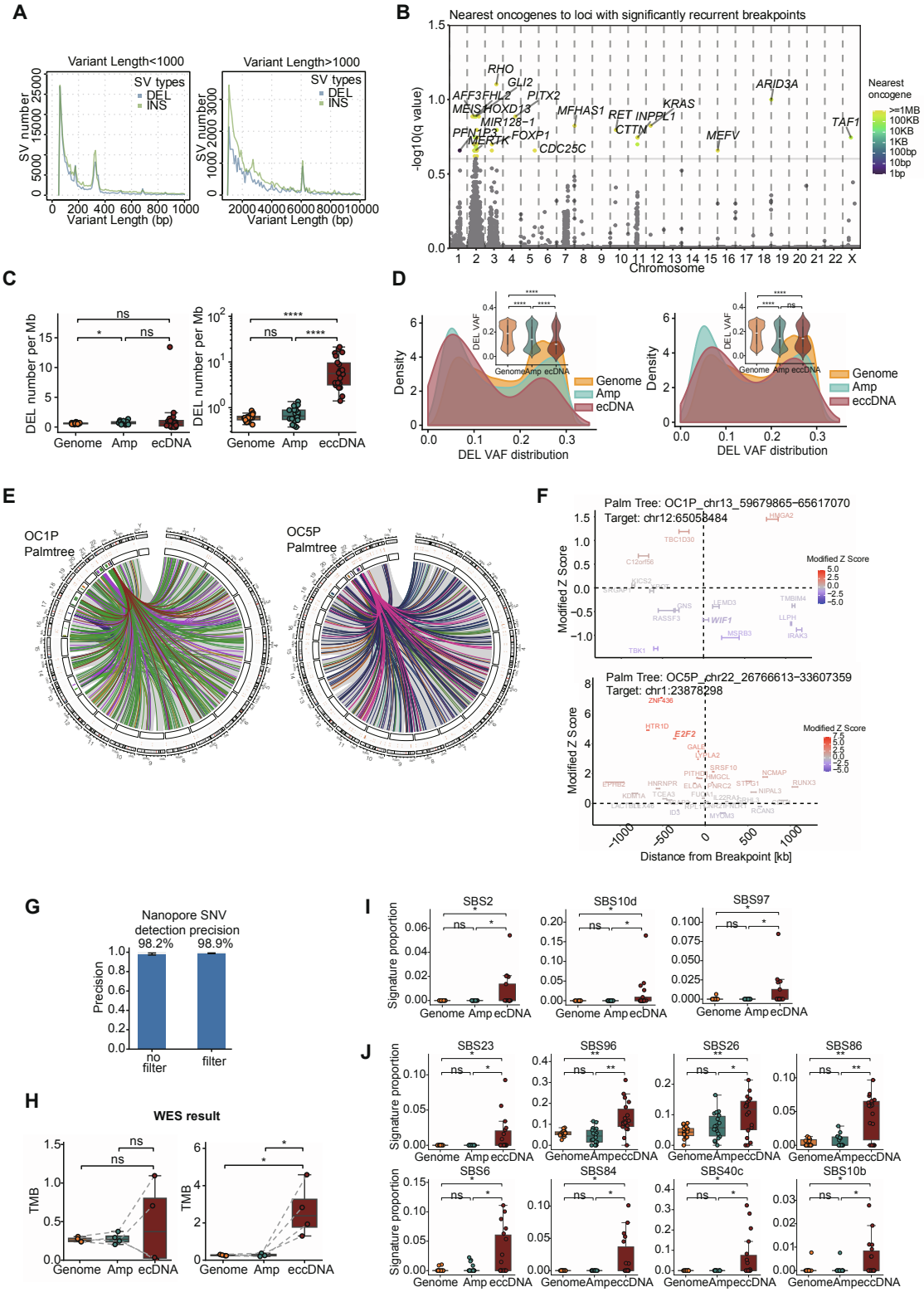

A
Variant Length<1000
Variant Length>1000
SV types
DEL
INS
SV number
Variant Length (bp)
B
Nearest oncogenes to loci with significantly recurrent breakpoints
-log10(q value)
Chromosome
RHO
GLI2
AFF3
FHL2
PITX2
MEIS1
HOXD13
MIR128-1
PFN1P3
FOXP1
MERTK
CDC25C
MFHAS1
RET
INPPL1
CTTN
KRAS
MEFV
ARID3A
TAF1
Nearest oncogene
>=1MB
100KB
10KB
1KB
100bp
10bp
1bp
C
DEL number per Mb
Genome
Amp
ecDNA
ns
*
****
D
Density
DEL VAF distribution
DEL VAF
Genome
Amp
ecDNA
E
OC1P
Palmtree
OC5P
Palmtree
F
Palm Tree: OC1P_chr13_59679865−65617070
Target: chr12:65058484
Palm Tree: OC5P_chr22_26766613−33607359
Target: chr1:23878298
Modified Z Score
Distance from Breakpoint [kb]
G
Nanopore SNV
detection precision
98.2%
98.9%
Precision
no filter
filter
H
WES result
TMB
I
SBS2
SBS10d
SBS97
Signature proportion
J
SBS23
SBS96
SBS26
SBS86
SBS6
SBS84
SBS40c
SBS10b

Figure S3. The SVs and mutational processes associated with circular DNA.

(A) Frequencies of different SV types with length less than 1 kb (left) and longer than 1 kb (right). The length of Alu is approximately 300 bp. The length of LINEs is approximately 6 kb.

(B) Manhattan plot shows the significantly recurrent mosaic SV breakpoints identified by FishHook and their distance to the nearest oncogenes in HGSOC. Each dot represents an FDR-adjusted (Benjamini-Hochberg) p-value of the distance, and a cutoff of 0.25 (horizontal solid line) was used to nominate significant hits.

(C) Comparison of the DEL number per Mb detected in ecDNA(left)/eccDNA(right), whole genomic regions and amplified regions in each tumor sample (two-sided Wilcoxon's signed rank-sum test, n=22, * P<0.05, **** P<0.0001).

(D) Comparison of the DEL VAF distribution detected in ecDNA(left)/eccDNA(right), whole genomic regions and amplified regions in each tumor sample (Mann-Whitney U test, n=22, **** P<0.0001). DEL: deletion. VAF: Variant Allele Frequency.

(E) Circos plot of interchromosomal rearrangements identified using mosaic SVs in HGSOC genomes, shown exemplarily.

(F) The modified z-scores for the expression of the genes affected by circle-derived rearrangements are shown for the representative genomic loci.

(G) Comparison of the accuracy of ONT and WES data in detecting the number of SNV mutations.

(H) Comparison of the SNV mutations per Mb detected in ecDNA/eccDNA from WES data, whole genomic regions and amplified regions in each tumor sample (paired Student's t-test, n=4, * P<0.05).

(I-J) Comparison of SBS signatures within genome, amplified regions and ecDNA(I)/eccDNA(J) regions. Only significant enriched SBS signatures are plotted (two-sided Wilcoxon rank sum test, * P<0.05, ** P<0.01).

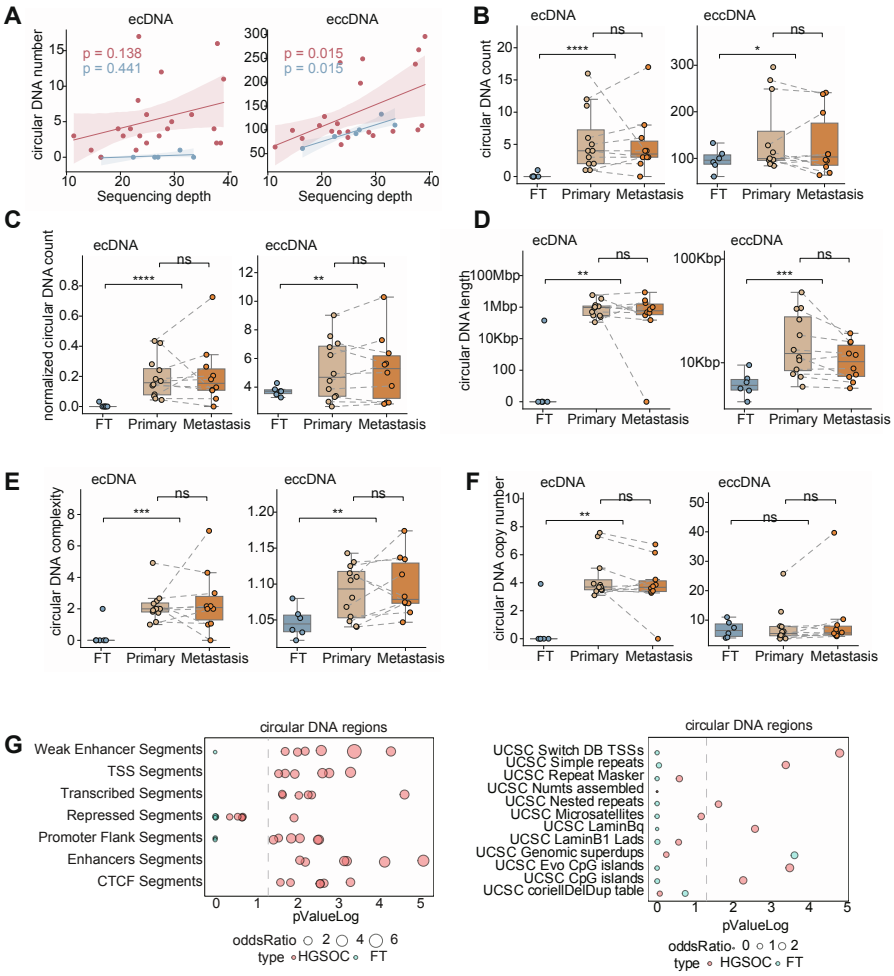


Figure S4. Distinct circular DNA properties.

(A) Significant positive correlation between eccDNA number and mapping depth of ONT data (Spearman's correlation, HGSOC is marked with red line; FT is marked with blue line) compared with no positive correlation for ecDNA.

(B) The ecDNA(left) and eccDNA(right) number for FT, primary and metastatic HGSOC samples (independent Student's t-test and paired Student's t-test separately, * P<0.05, **** P<0.0001).

(C) The normalized ecDNA(left) and eccDNA(right) number (normalized by mean coverage) for FT, primary and metastatic HGSOC samples (independent Student's t-test and paired t-test separately, ** P<0.01, **** P<0.0001).

(D-F) Length distribution, complexity and copy number of circular DNA (independent Student's t-test and paired Student's t-test separately, ** P<0.01, *** P<0.001).

(G) LOLA region set enrichment analysis for circular DNA regions in the LOLA Core database (encode_segmentation and ucsc_features). Region sets from normal FTs and HGSOC tissues with P < 0.05 are significant. P values were calculated using a two-sided Fisher's exact test.

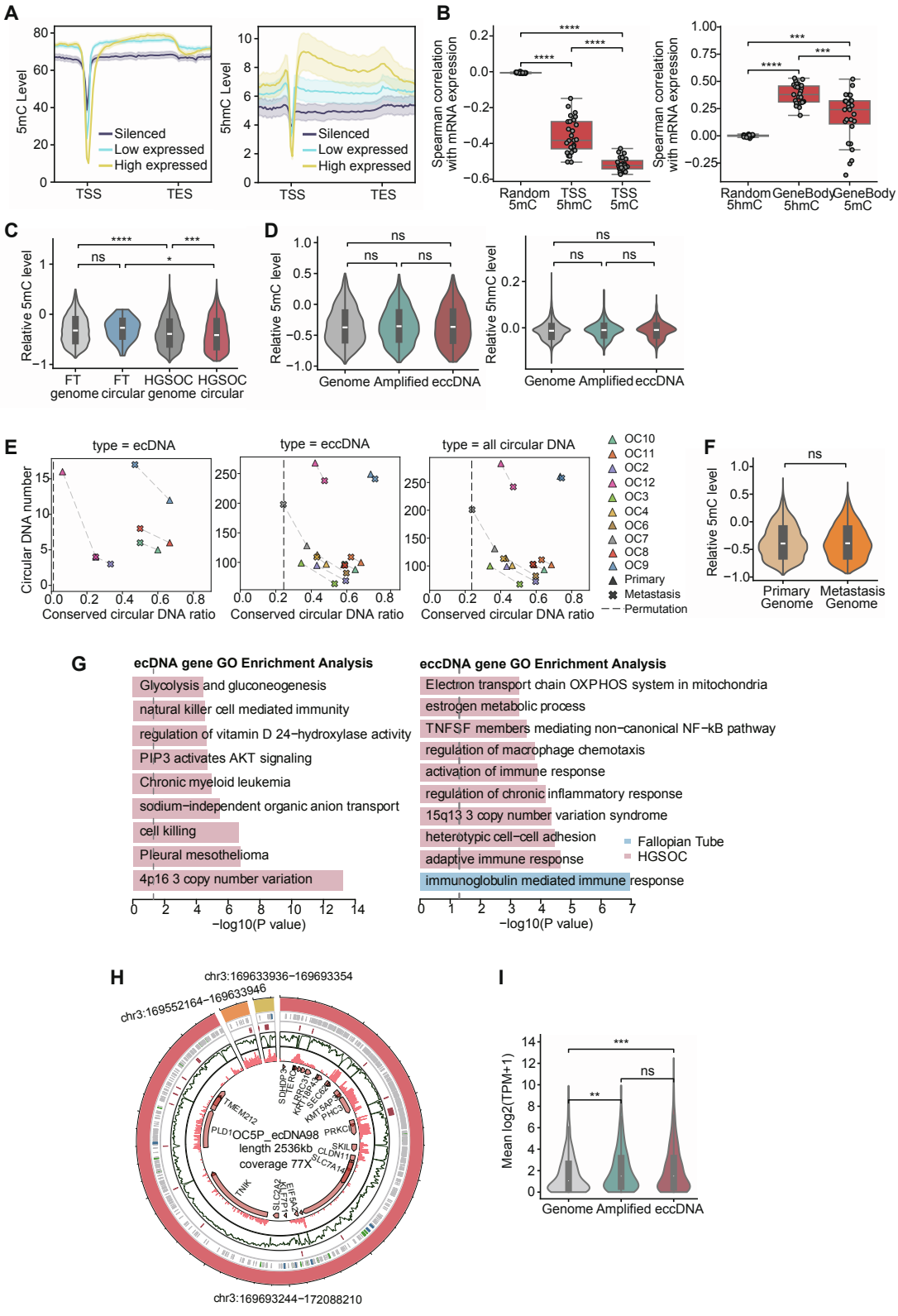


Figure S5. DNA methylation and gene features on circular DNA.

(A) 5mC and 5hmC profiles of genes expressed at high (yellow), low (blue), and silenced (purple) levels.

(B) Spearman's correlation of gene tss 5mC (5hmC) signals and gene body 5mC (5hmC) signals with

gene expression levels. Random regions were selected as controls (paired Student's t-test, *** P<0.001, **** P<0.0001).

(C) Relative 5mC levels between genomic and circular DNA regions for FT (*n*=6) and HGSOC (*n*=22) samples (Student's t-test, * P<0.05, *** P<0.001, **** P<0.0001).

(D) Relative 5mC and 5hmC levels between genomic, amplified and eccDNA regions for HGSOC samples (Student's t-test).

(E) The conserved circular DNA ratio derived from paired primary and metastatic tumor tissues from the same HGSOC patients. The vertical dashed line indicates the average percentage of conserved circular DNA ratio between any two samples.

(F) Relative 5mC levels on whole genomic regions for primary and metastatic samples (Student's t-test).

(G) Pathway enrichment analysis of genes on circular DNA in FT and HGSOC.

(H) Circos shows an example of several oncogenes located on the same ecDNA (OC5P samples).

(I) Gene expression comparing located on genomic, amplified and eccDNA regions for HGSOC samples (Student's t-test, ** P<0.01).

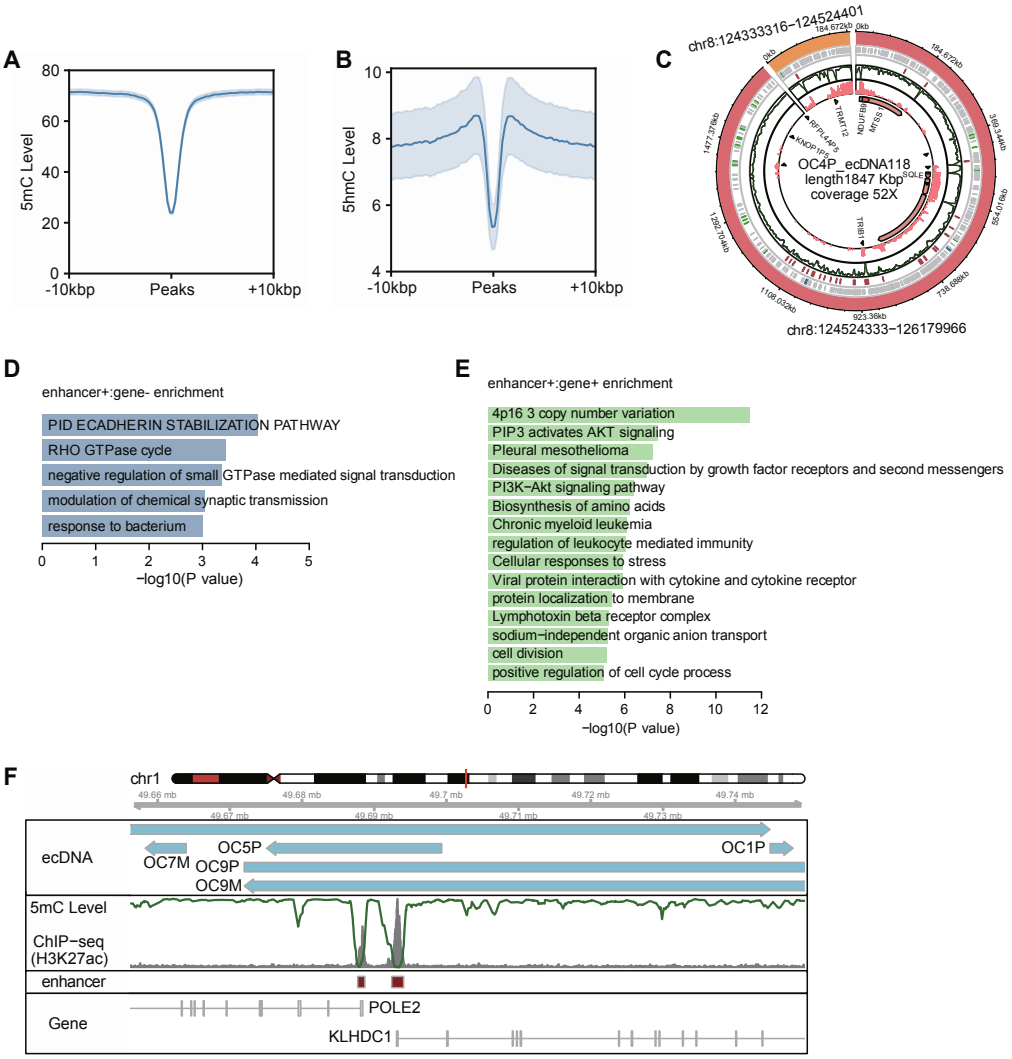


Figure S6. Enhancer characteristic on circular DNA.

(A-B) Distribution of 5mC and 5hmC level of enhancer regions(*n*=30,306) identified from public H3K27ac ChIP-seq (*n*=5). Mean was marked by solid line. Mean DNA methylation frequencies were calculated in 100-bp windows sliding every 10 bp.

(C) Circos shows an example of enhancer and gene co-amplification on the same ecDNA (OC4P samples).

(D-E) Pathway enrichment analysis of genes regulating by enhancer on circular DNA (left) and co-amplification on circular DNA (right).

(F) Methylation profile of the ecDNA regions near enhancers which may bind to tead1. Gene annotation and H3K27ac ChIP-seq levels of the region shown below.

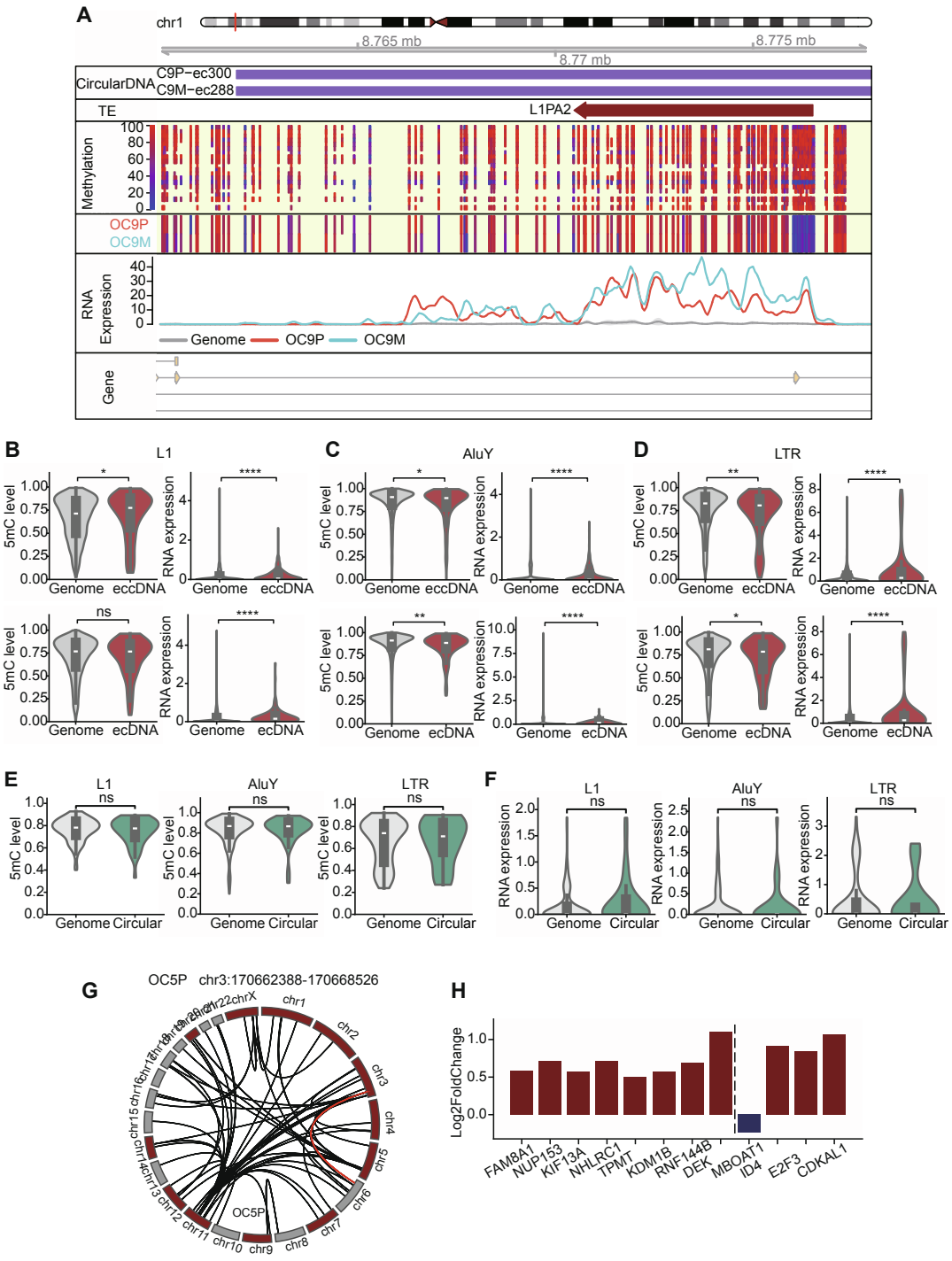


Figure S7. TE characteristic on circular DNA.

(A) The example of DNA methylation level and expression of LINE-1 on HGSOC circular DNA regions. (B-D) DNA methylation and RNA expression on TE subfamilies across linear genome and eccDNA/ecDNA for HGSOC samples (Mann-Whitney U test, * P<0.05, ** P<0.01, **** P<0.0001).

(E-F) DNA methylation and RNA expression on TE subfamilies across linear genome and circular DNA for normal FTs (Mann-Whitney U test). L1 length >1.2 kb, AluY length >280 bp, and LTR length >900 bp.

(G) Circos plot of LINE1 origin and insertion position, shown exemplarily.

(H) The modified z-scores for the expression of the genes affected by LINE1 insertion are shown for the representative genomic loci.

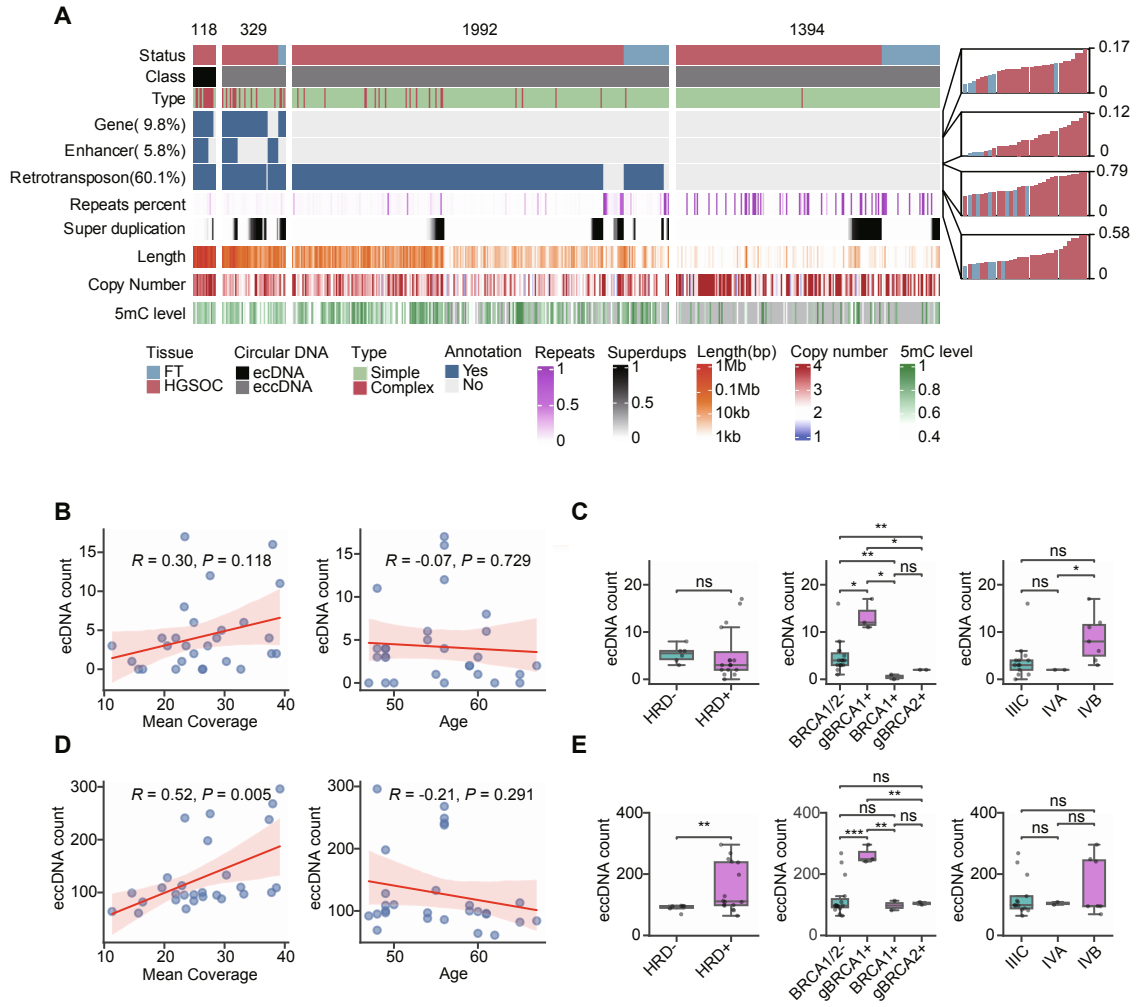


Figure S8. Heterogeneity of circular DNA profiles across HGSOC samples.

(A) The heatmap shows the distribution of all circular DNAs (n=3,833) containing gene, enhancer, transposon, super duplication, and repeat regions. Length distribution, copy number and 5mC levels of these circular DNAs are presented below. The right barplot indicates the relative circular DNA ratio of each sample.

(B) Correlations of sequencing depth (ONT) and patient age with ecDNA counts in HGSOC.

(C) Comparison of the ecDNA counts among HGSOC samples, including HRD status, BRCA status, and pathological stage (Welch's t-test, ns, not significant, * P<0.05, ** P<0.01).

(D) Correlations of sequencing depth (ONT) and patient age with eccDNA counts in HGSOC.

(E) Comparison of the eccDNA counts among HGSOC samples, including HRD status, BRCA status, and pathological stage (Welch's t-test, ** P<0.01, *** P<0.001).