Extrachromosomal Circular DNA Orchestrates Genome Heterogeneity and Outcome in Urothelial Bladder Carcinoma

Wei Lv[1,2]*, Xiaoguang Pan[2,3,4]*, Peng Han[2,3,4]*, Shuang Wu[5]*, Yuchen Zeng[6,7], Qingqing Wang[1], Lidong Guo[1,3], Mengyang Xu[3], Fei Wang[2,8], Zhe Xu[1,2], Yi Li[6], Huajing Teng[10], Conghui Li[2], Kairu Qi[3], Shang Gao[3], Tianxi Yu[5,9], Xin Cui[5,9], Chongjun Xiang[5,11], Haotian Tan[12], Yue Li[5,11], Ning Liang[5,9], Huiying Tao[5,11], Guohua Yu[13], Jia Mi[14], Fuyi Xu[14], Benjiao Gong[15], Lei Shi[5], Tao Wang[5], Lars Bolund[2,8,16,17], Huanming Yang[1,2,6], Wenting Wang[15‡], Hanbo Li[16‡], Jinrong Huang[2,16‡]*, Chunhua Lin[5‡], Yonglun Luo[2,6,8,16,17,‡]

Affiliations

[1]College of Life Sciences, University of Chinese Academy of Science, Beijing 100049, China.

[2]Lars Bolund Institute of Regenerative Medicine, Qingdao-Europe Advanced Institute for Life Sciences, BGI-Qingdao, Qingdao 266555, China.

[3]BGI Research-Qingdao, BGI, Qingdao, 266555, China.

[4]Department of Biology, University of Copenhagen, Copenhagen 2100, Denmark.

[5]Department of Urology, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, Yantai, Shandong, 264000, China.

[6]IBMC-BGI Center, the Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China.

[7]College of Life Sciences, Tianjin University, Tianjin 300072, China.

[8]Department of Biomedicine, Aarhus University, Aarhus, Denmark.

[9]School of Clinical Medicine, Weifang Medical University, Weifang, 261042, China.

[10]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital & Institute, Beijing, China.

[11]The 2nd Medical College of Binzhou Medical University, Yantai, Shandong, 264003,China.

[12]Department of Urology, China-Japan Friendship Hospital (Institute of Clinical Medical Sciences), Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, 100029, China.

[13]Department of Pathology, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, Yantai, Shandong, 264000, China.

[14]Precision Medicine Research Centre, School of Pharmacy, Binzhou Medical University. Yantai, China, 264003.
[15]Department of Central Laboratory, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, Yantai, Shandong 264000, China.
[16]BGI-Shenzhen, Shenzhen 518083, China.
[17]Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark.

*These authors contributed equally: Wei Lv, Xiaoguang Pan, Peng Han, Jinrong Huang, and Shuang Wu
‡Corresponding should be addressed to Yonglun Luo (alun@biomed.au.dk), Chunhua Lin (chunhua.lin@qdu.edu.cn), Jinrong Huang (huangjinrong@genomics.cn), Hanbo Li (lihanbo@genomics.cn), Wenting Wang (wwting_78@163.com)

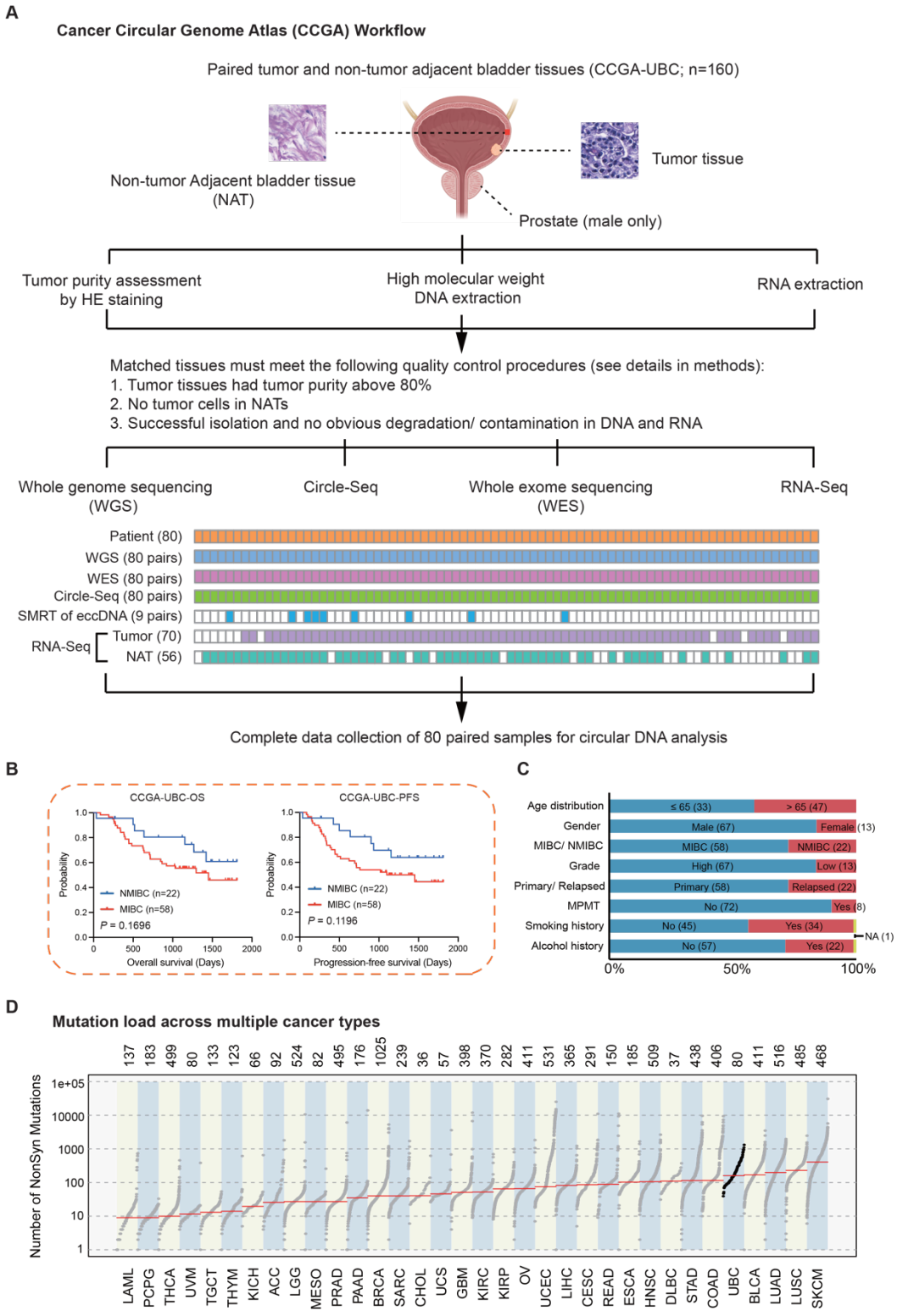


Figure S1. The Experimental Workflow of the CCGA-UBC Circular DNA Study and The Clinical Characteristics of CCGA-UBC Patients, Related to the Figure 1 and Star methods. (A) A total of 80 pairs of freshly snap-frozen tumor tissues and NATs were collected from UBC patients. Hematoxylin-Eosin (HE) staining was performed to assess the tumor purities. Sample filtering was performed based on the following criteria: 1) Tumor tissues had tumor purity above 80%; 2)No tumor cells were observed in NATs; 3) Successful isolation of high molecular

weight genomic DNA (HMW gDNA) and qualified RNA (RNA integrity number > 5) from both tumor and NAT samples. Finally, WGS, WES, Circle-Seq, and RNA-Seq were performed on each qualified sample. SMRT of eccDNA was performed on 9 pairs of tumor-NAT samples. RNA samples from 10 tumor tissues and 24 NATs were excluded due to low quality. (B) Survival analysis of patients with MIBC and NMIBC (log-rank test). (C) Summary of clinical characteristics of 80 CCGA-UBC patients. Clinical features include age distribution, gender, MIBC/ NMIBC, grade, primary/ relapsed, MPMT (multiple primary malignant tumors), smoking history, and alcohol history. (D) Pan-cancer distribution of the number of non-synonymous mutations. Each dot represents a tumor sample from the TCGA datasets or CCGA-UBC cohort. Red bars reflect the median number of non-synonymous mutations for each cancer type.

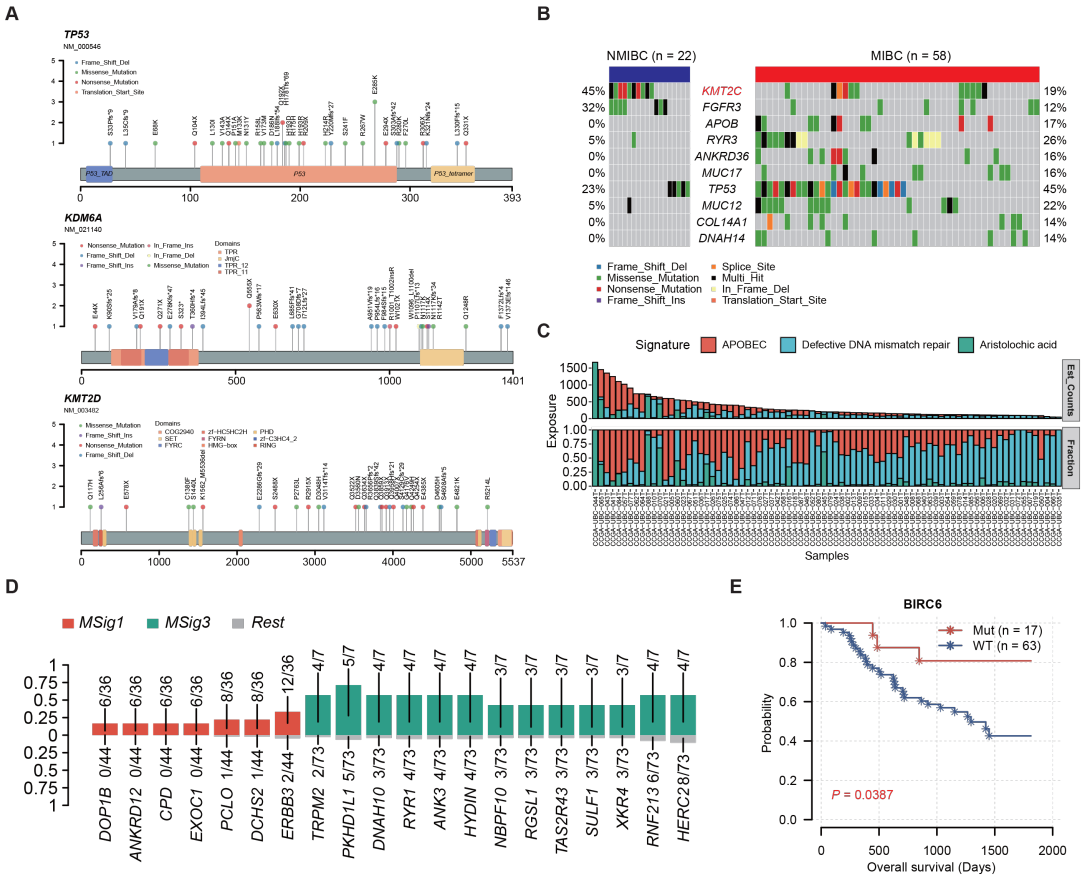


Figure S2. Somatic Mutation Spectrum of CCGA-UBC Samples, Related to Figure 1. (A) Somatic mutations identified in the *TP53*, *KDM6A*, and *KMT2D* proteins. (B) Gene alterations in NMIBC and MIBC. Genes with significant differences in mutation frequency between NMIBC and MIBC (Fisher's exact test, $P < 0.05$) are marked with red. (C) The mutation number of each mutational signature (Est_Counts) and normalized contributions across samples (Fraction). (D) Genes with significantly more frequent mutations in MSig1 or MSig3, compared to the rest of the cases (Fisher's exact test, $P < 0.01$). (E) Survival analysis of patients with *BIRC6* mutation versus *BIRC6* wild type (log-rank test).

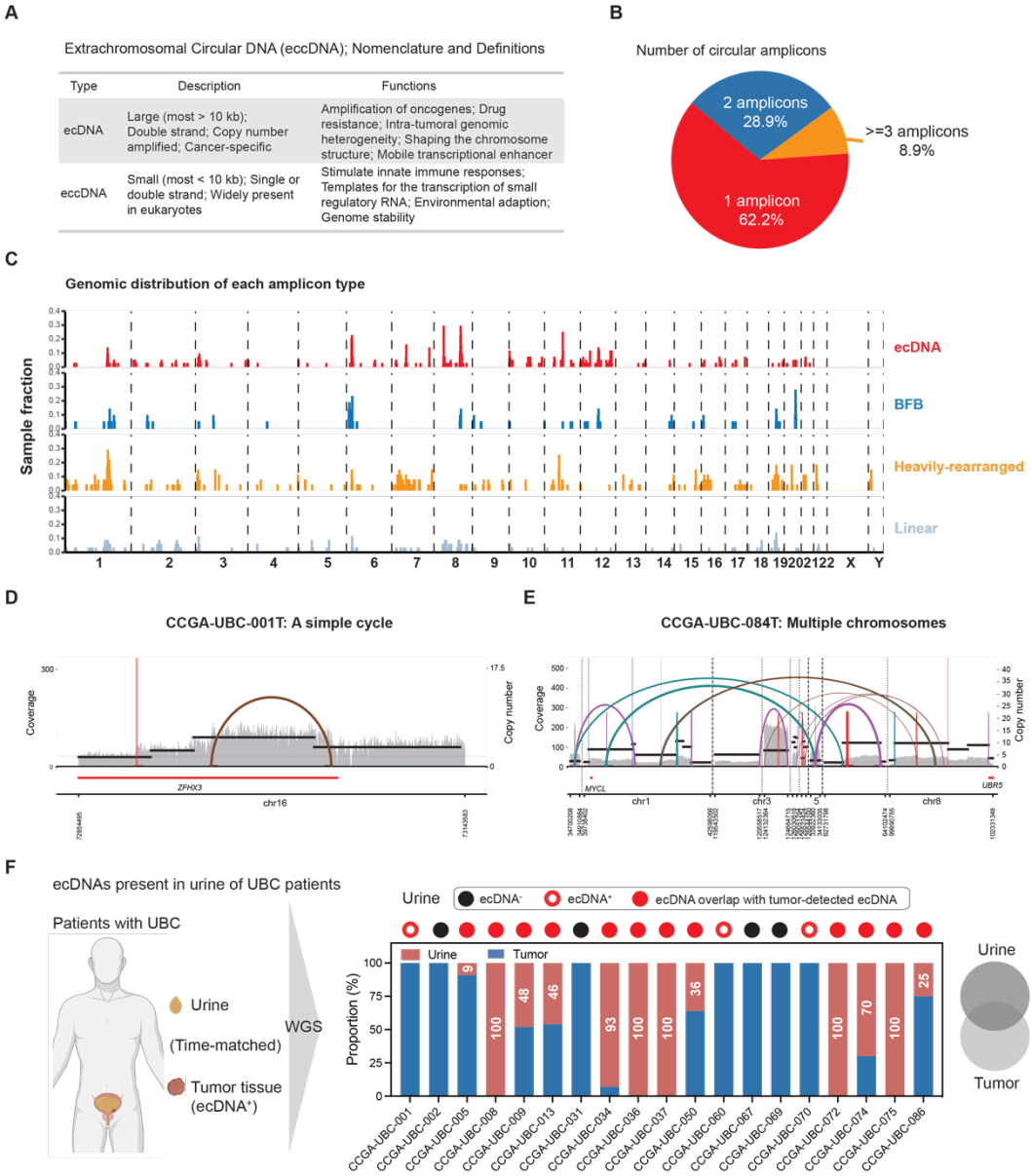


Figure S3. Focal Amplification Landscape of CCGA-UBC Samples, Related to Figure 2. (A) Summary of the definition and functions of two main types of extrachromosomal circular DNA. (B) Pie chart showing the percentages of tumors with indicated circular amplicon numbers among the total 45 ecDNA$^+$ cases. (C) Genomic distribution of each amplicon class, including ecDNA, BFB, Heavily-rearranged, and Linear. (D-E) AA-generated structural variant (SV) and breakpoint graph of examples of a simple cycle (D) and a complex multi-chromosomal circular amplification (E). (F) ecDNAs present in the urine sample. Left panel: the workflow for ecDNA detection in urine samples. Time-matched urine samples from patients carrying ecDNA$^+$ tumors (n =19) were performed with WGS. Right panel: Overlap of ecDNA detected in tumors and corresponding urine samples.

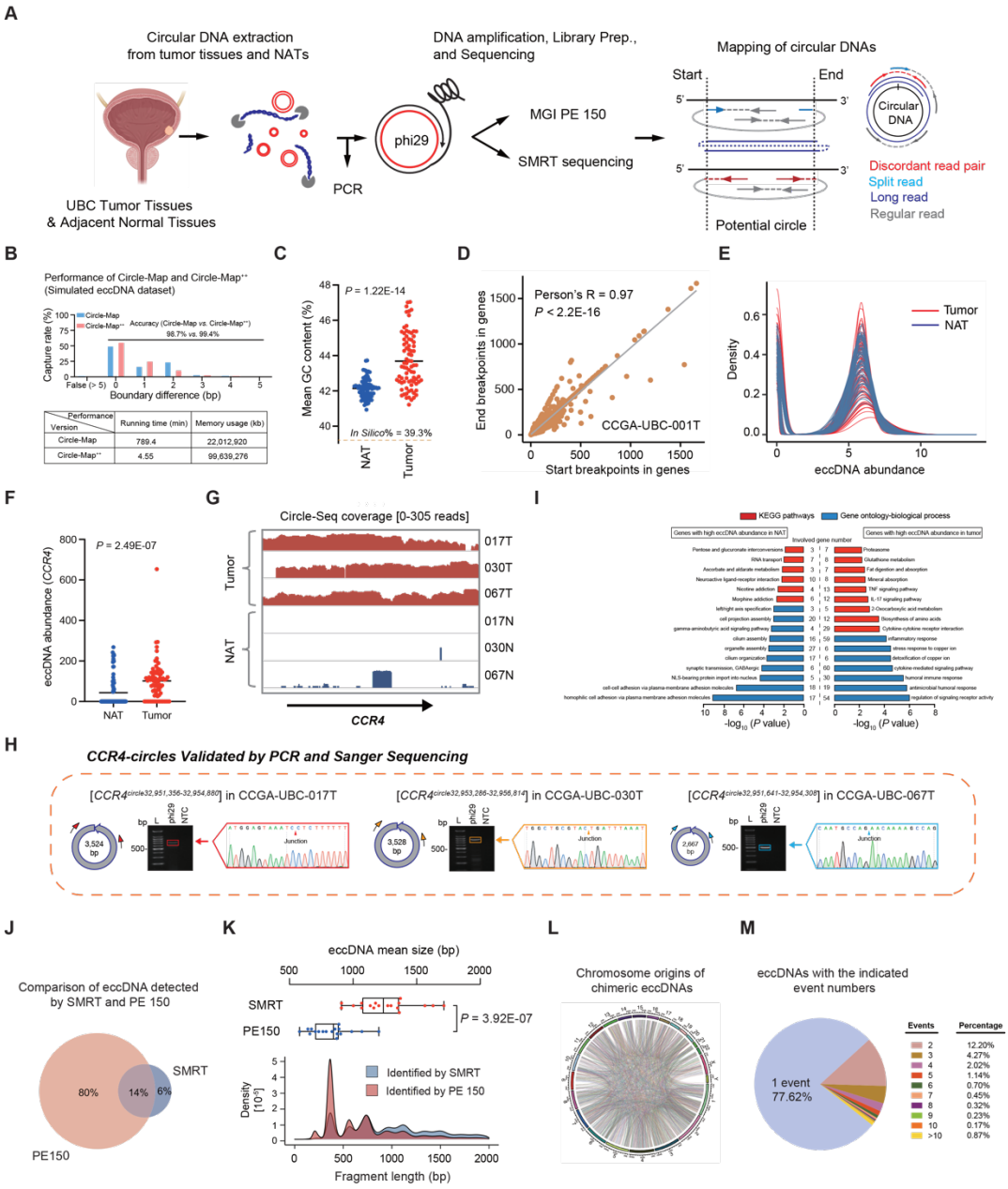


Figure S4. Genome-scale Map of Extrachromosomal Circular DNAs in CCGA-UBC Samples, Related to STAR Methods and Figure 3

(A) Circle-Seq method for mapping of circular DNAs. Paired normal-tumor samples were collected from 80 CCGA-UBC patients. Total high molecular weight genomic DNA (HMW *g*DNA) was isolated from tissues using a magnetic-based approach. The linear portions of HMW *g*DNA were hydrolyzed by Exonuclease V (Exo V). PCR was performed to validate the removal of linear DNA. The purified circular DNAs were amplified by phi29 polymerase. The amplified products were subjected to MGI PE150 sequencing and long-read SMPT sequencing (PacBio). The eccDNAs were detected from sequencing data by Circle-Map^{++} and the Consensus eccDNA generation method (see methods). (B) Evaluation of the performance of Circle-Map and Circle-Map^{++} in the detection of eccDNAs. Circle-Map and Circle-Map^{++} were evaluated on a simulated eccDNA dataset. Shown are the capture rate and accuracy of Circle-Map and Circle-Map^{++} (top) and the running time and memory usage of Circle-Map and Circle-Map^{++} (bottom). (C) Comparison of the mean GC content between CCGA-UBC tumors and NATs (Paired t-

test). (D) Correlation of the number of start breakpoints and the number of end breakpoints in each gene (CCGA-UBC-001T) (Person correlation test). The start and end breakpoints of each eccDNA refer to the upstream and downstream boundary of the circularized genomic region. (E) Density plot showing the eccDNA abundance distribution in NATs (blue) and tumors (red). A non-normally distribution was observed. (F) Comparison of the eccDNA abundance of *CCR4* gene between tumors and NATs (Paired t-test). (G) Circle-Seq reads coverage across the *CCR4* locus on chromosome 3 in three example cases of paired tumors and NATs. (H) Validation of three typical *CCR4*-related circles by gel and Sanger sequencing. (I) Pathways enriched for genes with differential eccDNA abundance between tumors and NATs. (J) Venn diagram showing the overlap of non- chimeric eccDNAs identified by SMRT and PE150. (K) Top panel: comparison of the mean length of eccDNA detected by SMRT and PE150 (Paired t-test). Bottom panel: Fragment length distribution of eccDNA detected by SMRT and PE 150 (Polled data from 18 cases from each group). (L) Circles plot presenting the chromosome origins of all identified chimeric eccDNAs. (M) Pie chart displaying the percentage eccDNA with indicated event numbers. The eccDNAs were detected by SMRT.

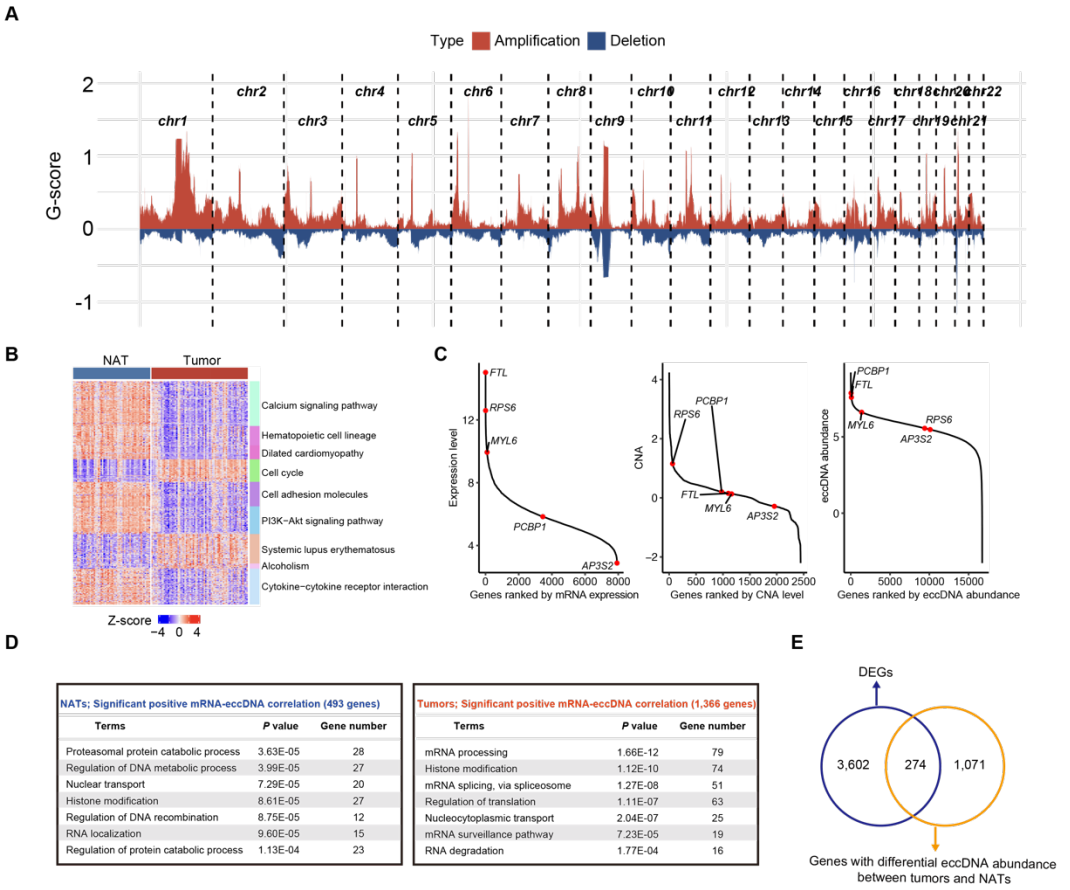
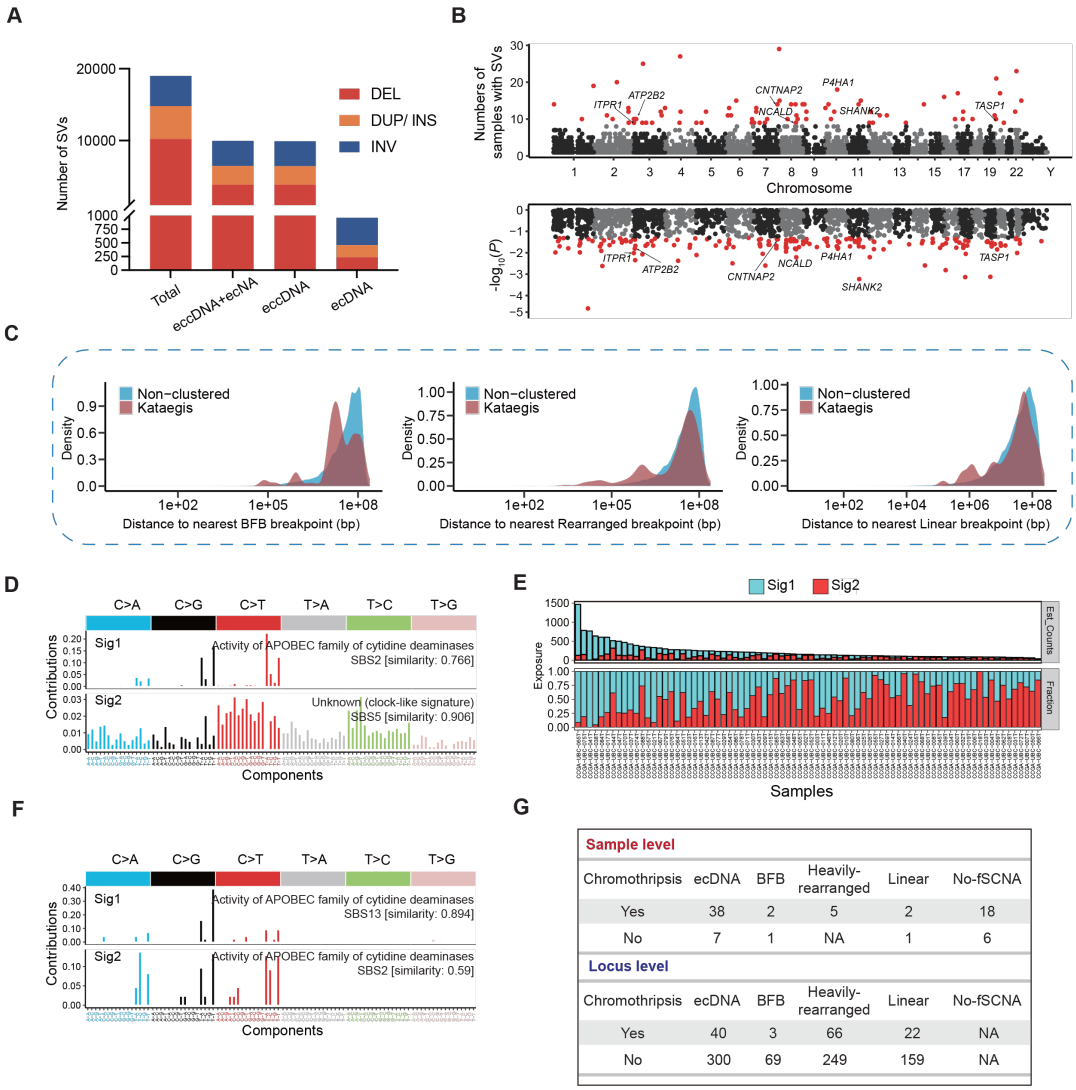


Figure S5. Characteristics of mRNA and Copy Number Alterations, and Correlations of mRNA to eccDNA in CCGA-UBC Samples, Related to Figure 4

(A) Genomic regions with somatic focal amplifications and deletions. Red, amplifications; Blue, deletions. (B) Differentially expressed genes (DEGs) among CCGA-UBC tumors and NATs and their related biological pathways (absolute $\log_2$ fold change > 1; $P < 0.01$). (C) Ranked mRNA expression, eccDNA abundance, and CNA value in CCGA-UBC-065 tumor samples. (D) Pathways enriched in genes with significantly eccDNA-mRNA correlation in NATs and tumors. (E) Overlap of genes with a significant change in both gene expression levels and eccDNA abundance in tumors and NATs.



Sample level

Chromothripsis	ecDNA	BFB	Heavily-rearranged	Linear	No-fSCNA
Yes	38	2	5	2	18
No	7	1	NA	1	6

Locus level

Chromothripsis	ecDNA	BFB	Heavily-rearranged	Linear	No-fSCNA
Yes	40	3	66	22	NA
No	300	69	249	159	NA

Figure S6. Frequent Co-localization of Circular DNA with other Genomic Alterations, Related to Figure 5

(A) Number of SVs overlapping ecDNA and eccDNA. Abbreviations: DEL, deletion; DUP: duplication; INS: insertion; INV: inversion. (B) Top: Number of samples with SV breakpoints in each gene. Genes affected by SVs in ⩾8 samples are marked with red. Bottom: Association of eccDNA abundance with SVs. eccDNA abundance was compared for the sample with and without SVs of each gene. Genes with significant differential eccDNA abundance are marked with red. (C)Distance to the nearest BFB (Left), Rearrangement (Middle), and Linear (Right). breakpoints for kataegis mutations and non-clustered mutations. (D) Mutational signatures of all kataegis. (E) The mutation number of each mutational signature (Est_Counts) and normalized contributions across samples (Fraction). (F) Mutational signatures of kataegis within 10 kb of ecDNA breakpoints. (G) Co-occurrence of chromothripsis with focal amplification. Sample level, chromothripsis distribution in different amplification samples. Locus level, the overlap of chromothripsis with each amplicon type.

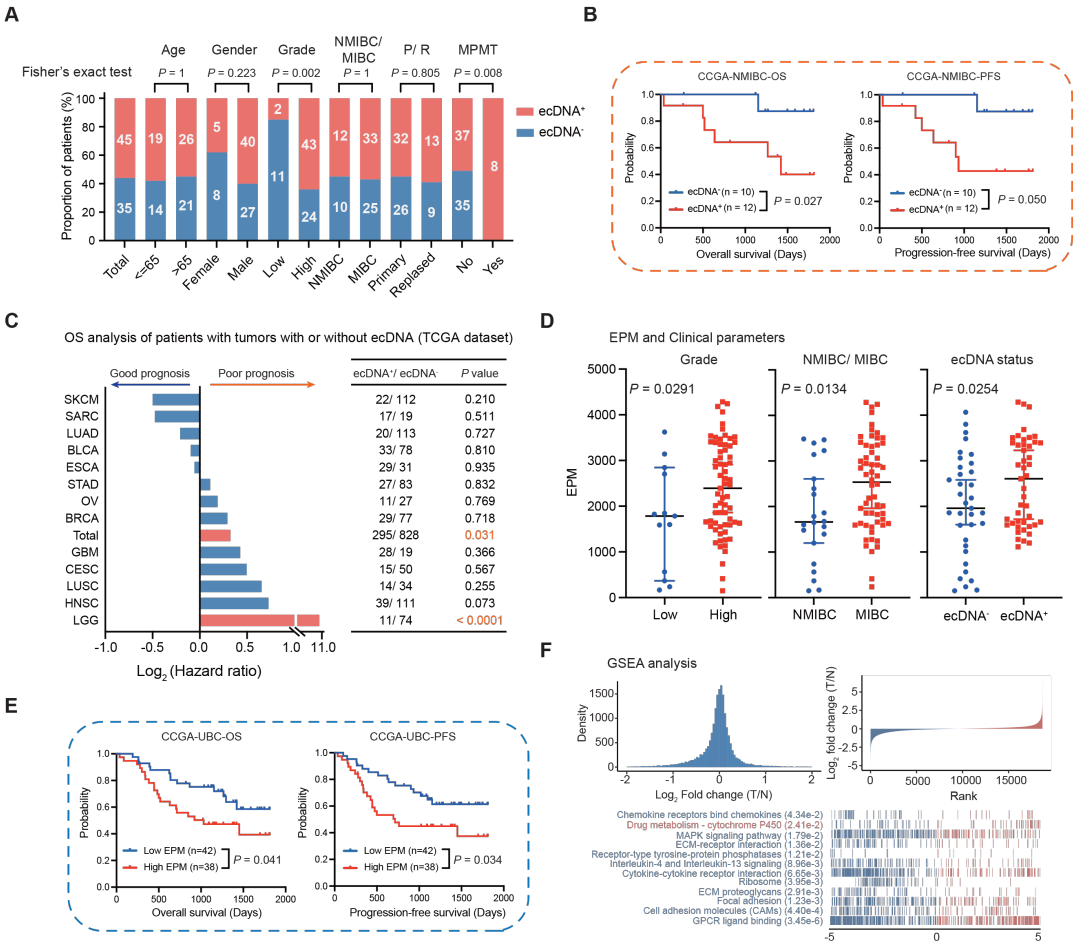


Figure S7 Mutation load between ecDNA$^-$ and ecDNA$^+$ tumors and relationship between circular DNAs and clinical variables, Related to Figure 6 and Figure 7

(A) Comparison of TMB between ecDNA$^-$ and ecDNA$^+$ tumors in 13 cancer types from TCGA datasets. (B-C) Lollipop plots of *FGFR3* (B) and *TTN* (C) show the locations of somatic mutations in the tumors with and without ecDNA. (D) Survival analysis of CCGA-MIBC patients with and without ecDNA (log-rank test). (E) The association of EPM with 9 clinical and molecular variables (Wilcoxon rank-sum test).

SUPPLEMENTARY TABLES

Table S1. Clinicopathologic Information and Multi-omics Data in CCGA-UBC cohort. Related to Figure 1

Table S2. Detailed Focal Amplification Information in CCGA-UBC cohort. Related to Figure 2

Table S3. Comprehensive eccDNA Profiling of CCGA-UBC samples. Related to Figure 3

Table S4. Characteristics of mRNA and Copy Number Alterations, and Correlations of mRNA to eccDNA in CCGA-UBC Samples. Related to Figure 4

Table S5. SV, Cluster Mutation, and Chromothripsis Data in CCGA-UBC cohort. Related to Figure 5

Table S6. Genes correlated with Circular DNA Formation in UBC. Related to Figure 6