RESEARCH ARTICLE

Spatial-Temporal Diversity of Extrachromosomal DNA Shapes Urothelial Carcinoma Evolution and the Tumor Immune Microenvironment

Wei Lv^{1,2,3}, Yuchen Zeng¹, Conghui Li⁴, Yuan Liang⁵, Huiying Tao⁶, Yanfen Zhu⁷, Xiaolong Sui⁶, Yue Li⁶, Shiqi Jiang¹, Qingqing Gao¹, Elias Rodriguez-Fos^{8,9}, Gino Prasad¹⁰, Yuanmei Wang¹, Run Zhou¹, Zhe Xu¹¹, Xiaoguang Pan^{4,11}, Linlin Chen¹, Xi Xiang¹², Huajing Teng¹³, Chaoyang Sun^{14,15}, Tianyu Qin^{14,15}, Wei Dong¹, Yongwei Li⁶, Xun Lan¹⁶, Xuesong Li¹⁷, Lin Lin³, Lars Bolund^{1,3,11}, Huanming Yang¹, Roel G.W. Verhaak¹⁸, Bishoy M. Faltas^{19,20,21,22}, Jacob B. Hansen⁴, Sihan Wu²³, Paul S. Mischel^{24,25}, Anton G. Henssen^{8,9,26,27}, Vineet Bafna^{10,28}, Jens Luebeck¹⁰, Birgitte Regenberg⁴, Yonglun Luo³, Chunhua Lin⁶, and Peng Han^{11,12}

ABSTRACT

Extrachromosomal DNA (ecDNA) presents a promising target for cancer therapy; however, its spatial-temporal diversity and influence on tumor evolution and

the immune microenvironment remain largely unclear. We apply computational methods to analyze ecDNA from whole-genome sequencing data of 595 patients with urothelial carcinoma. We demonstrate that ecDNA drives clonal evolution through structural rearrangements during malignant transformation and recurrence of urothelial carcinoma. This supports a model wherein tumors evolve via the selective expansion of ecDNA-bearing cells. Through multiregional sampling of tumors, we demonstrate that ecDNA contributes to the evolution of multifocality and increased intratumoral heterogeneity. ecDNA is present in 36% of urothelial carcinoma tumors and correlates with an immunosuppressive phenotype and poor prognosis. Single-cell RNA sequencing analyses reveal that ecDNA⁺ malignant cells exhibit diminished expression of MHC class I molecules, enabling them to evade T-cell immunity. Finally, we show that sequencing of urinary sediment-derived DNA has excellent specificity in detecting ecDNA.

SIGNIFICANCE: Our comprehensive analysis of ecDNA in urothelial carcinoma reveals its crucial role in driving the evolution and heterogeneity of multifocal cancer, as well as its early involvement in tumorigenesis. Moreover, this study sheds light on immune evasion mechanisms associated with ecDNA and offers valuable insights for developing targeted therapeutic strategies.

INTRODUCTION

Urothelial carcinoma is a prevalent malignancy that originates from the transitional epithelium of the bladder, ureter, renal pelvis, and urethra. Urothelial carcinoma often presents as multifocal lesions, indicative of its heterogeneous nature. The progression of urothelial carcinoma is a complex, multistage tumorigenesis process, typically unfolding through two pathways: (i) from urothelial hyperplasia to

New York, New York. ²¹Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, New York. ²²Department of Cell and Developmental Biology, Weill Cornell Medicine, New York, New York. ²³Children's Medical Center Research Institute, University of Texas Southwestern Medical Center, Dallas, Texas. ²⁴Department of Pathology, Stanford University School of Medicine, Stanford, California. ²⁵Sarafan Chemistry, Engineering, and Medicine for Human Health (Sarafan ChEM-H), Stanford University, Stanford, California. ²⁶Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany. ²⁷German Cancer Consortium, partner site Berlin, and German Cancer Research Center, Heidelberg, Germany. ²⁸Halıcıoğlu Data Science Institute, University of California at San Diego, San Diego, California.

W. Lv, Y. Zeng, C. Li, and Y. Liang share co-first authorship of this article.

Corresponding Authors: Wei Lv, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, No. 3 Qingchun East Road, Shangcheng District, Hangzhou 310016, Zhejiang, China. E-mail: wei_lv2024@163.com; Birgitte Regenberg, Section for Ecology and Evolution, Universitetsparken 13, Copenhagen Ø DK-2100, Denmark. E-mail: bregenberg@bio.ku.dk; Yonglun Luo, Department of Biomedicine, Building 1116-152, Høegh-Guldbergs Gade 10, Aarhus 8000, Denmark. E-mail: alun@biomed.au.dk; Chunhua Lin, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, No. 20, Yuhuangding East Road, Zhifu District, Yantai 264000, Shandong, China. E-mail: chunhua.lin@qdu.edu.cn; and Peng Han, The 7th Affiliated Hospital of Sun Yat-sen University, 628 Zhenyuan Road, Xinhu Street, Guangming District, Shenzhen 518000, Guangdong, China. E-mail: ucphpeng@gmail.com Cancer Discov 2025;XX:1-22

doi: 10.1158/2159-8290.CD-24-1532

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

© 2025 The Authors; Published by the American Association for Cancer Research

¹HIM-BGI Omics Center, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, China. ²Department of Urology & Andrology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China. ³Department of Biomedicine, Aarhus University, Aarhus, Denmark. ⁴Department of Biology, University of Copenhagen, Copenhagen, Denmark. ⁵CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing, China. 6Department of Urology, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, Yantai, China. ⁷The Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, China. ⁸Department of Pediatric Oncology and Hematology, Charité - Universitätsmedizin Berlin, Humboldt-Universität zu Berlin, Berlin, Germany. ⁹Experimental and Clinical Research Center of the MDC and Charité Berlin, Berlin, Germany. ¹⁰Department of Computer Science and Engineering, University of California at San Diego, San Diego, California. ¹¹Lars Bolund Institute of Regenerative Medicine, Qingdao-Europe Advanced Institute for Life Sciences, Qingdao, China. ¹²Scientific Research Center, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen, China. ¹³Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital & Institute, Beijing, China. ¹⁴Department of Gynecological Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ¹⁵National Clinical Research Center for Obstetrics and Gynecology, Cancer Biology Research Center (Key Laboratory of the Ministry of Education), Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ¹⁶Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing, China. ¹⁷Department of Urology, Peking University First Hospital, Beijing, China. ¹⁸Department of Neurosurgery, Yale School of Medicine, New Haven, Connecticut. ¹⁹Department of Medicine, Weill Cornell Medicine, New York, New York. ²⁰Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine,

papillary urothelial carcinoma and eventually to invasive cancer (1–3); and (ii) from urothelial dysplasia to carcinoma *in situ* and ultimately to invasive cancer. The pathogenesis of urothelial carcinoma involves a multifaceted interplay between genetic predisposition and environmental factors. Although considerable efforts focus on researching optimal treatment approaches, elucidating the molecular mechanisms driving malignant transformation and progression is crucial to optimize and personalize treatment (4, 5).

Oncogene copy-number amplification (CNA) plays a crucial role in urothelial oncogenesis (6-10). These amplifications often manifest focally, with high amplitude, and are frequently present in the form of extrachromosomal DNA (ecDNA) circles (11, 12). Also known as double minutes (13), ecDNA refers to circular DNA particles derived from chromosomes, typically ranging from 100 kilobases (kbp) to several megabases (Mbp) in length (14, 15). ecDNAs have been found in multiple cancers (11, 16-18), and their frequency and composition are significantly influenced by the tissue of origin (19, 20). ecDNA can rapidly accumulate in cancer cells given uneven mitotic segregation of ecDNA attributed to their lack of centromeres (21). This inheritance trait allows for ultra-rapid evolution of gene copy number (CN) and furnishes cancer cells with a competitive advantage over noncancerous cells when confronting selective pressures in cytotoxic treatments and tumor ecosystems (21-23). Indeed, the presence of ecDNA is associated with an aggressive phenotype and unfavorable prognosis in several tumor types (11, 18–20, 24). We and others have previously demonstrated that ecDNA shapes genetic intratumoral heterogeneity (10, 18, 25-28). Despite these advances, it is still largely unknown when ecDNA is acquired in the progression of many cancers such as urothelial carcinoma and how the dynamics of ecDNA in tumor space and time affect urothelial carcinoma tumor heterogeneity. One possibility is that ecDNA can emerge later in urothelial carcinoma development and evolution, potentially reshaping the tumor immune microenvironment (TIME) in urothelial carcinoma, because (i) the acquisition of CNAs occurs late in clonal expansion in the urothelium (29); and (ii) the immunomodulatory genes are frequently encoded on ecDNA (19, 30). Alternatively, the nonchromosomal inheritance of ecDNA could arise early in tumor development, as has been shown for the transition from Barrett's esophagus to cancer, with the intratumoral genetic heterogeneity arising as a consequence of random segregation (21, 30).

Therefore, to delineate the molecular evolution of ecDNA in urothelial carcinoma pathogenesis, we analyzed singleregional and multi-regional whole-genome sequencing (WGS), whole-exome sequencing (WES), and Circle-seq on a large panel of tumors from 595 patients prospectively recruited into the Chinese Urothelial Carcinoma Genome Atlas (CUGA) study. To understand the transcriptional effects of ecDNA, we also analyzed bulk RNA sequencing (RNA-seq) and single-cell RNA-seq (scRNA-seq) data from the tumors. We show that urinary ecDNA can be a valuable marker for the detection of urothelial carcinoma in liquid biopsies.

RESULTS

ecDNA Prevalence and Features in 1,411 Urothelial Carcinoma Whole Genomes

To explore the ecDNA landscape in urothelial carcinoma, we generated a large-scale, uniformly processed WGS inventory. This inventory included 753 tumor samples, 296 tissues adjacent to tumor tissues [adjacent tissues (AT)], 195 blood samples, and 167 preoperative (1-3 days before definitive surgery) voided urine samples. Samples originated from both the Yantai Yuhuangding Hospital UC dataset (YHD-UC, n = 493patients) and the Beijing Institute of Genomics (BIG) dataset (BIG-UC, n = 102 patients; Fig. 1A; refs. 31–33). This combined cohort (CUGA) of 595 patients represents a heterogeneous group of urothelial carcinomas in terms of age, pathologic stage and grade, tumor location, and molecular subgroups (Supplementary Tables S1-S8). None of the patients received any anticancer treatment before the collection of tumor specimens during surgery. Notably, WGS data of 10% (61 of 595) of the patients involved multiregional tumors (57 cases) or matched primary-recurrent tumors (4 cases). We applied the AmpliconArchitect and AmpliconClassifier pipelines (34) to identify and reconstruct ecDNA from biopsy WGS data (Fig. 1A). Previous reports indicated that AmpliconArchitect and AmpliconClassifier have high efficiency in ecDNA detection, with a sensitivity of 90% (30).

We first examined the prevalence of ecDNA in our cohort. ecDNAs were found in 36% of tumor samples (217 of 595). Notably, ecDNAs were also present in 23% of urine samples (39 of 167) and 3% of ATs (9 of 296; Fig. 1B; Supplementary Tables S9 and S10). ecDNA was detected in a similar frequency across the BIG-UC (29%) and YHD-UC cohorts (38%; Supplementary Fig. S1A and S1B). Moreover, the prevalence of ecDNA in urothelial carcinoma exceeded that of most cancer types previously reported in The Cancer Genome Atlas and Pan-Cancer Analysis of Whole Genomes cohorts, including hepatocellular carcinoma (10%) and gastric cancer (9%; Supplementary Fig. S1A; Supplementary Table S11), suggesting that ecDNA may play an important role in urothelial carcinoma development.

To understand the function of ecDNA in urothelial carcinoma development, we explored its structural and sequence features. We observed significant structural variability in ecDNA, including single-locus ecDNA (35%, 113 of 440) and ecDNA formed from two or more genomic segments (65%, 287 of 440; Supplementary Fig. S2A). Whereas ecDNA and other focal somatic CNAs (fSCNA) exhibit significant overlap in their genomic distribution [Supplementary Fig. S2B; iSTAT test (35), P = 0.0001], ecDNA distinguished itself through markedly higher structural complexity, larger amplicon size, and increased CN (Supplementary Fig. S2C; Supplementary Table S12; Wilcoxon rank-sum test; complexity score, *P* = 2.39e–30; amplicon size, *P* = 4.71e–17; CN, *P* = 1.23e–87). Localized hypermutation, or kataegis, is known to be prevalent in ecDNAs (36, 37). We found that kataegic mutation clusters were significantly more frequent in ecDNAs than in other fSCNAs (Supplementary Fig. S2D; two-sided Fisher exact test, P = 3.023e-07). These findings highlight distinct features by which ecDNA shapes genetic heterogeneity. Furthermore, nearly half (48%, 105 of 217) of

RESEARCH ARTICLE



Figure 1. ecDNA can arise in flat urothelial lesions. **A**, Study samples and ecDNA detection. A total of 753 tumor samples, 296 adjacent tumor tissues (AT), 195 blood samples, and 167 preoperative (1-3 days before definitive surgery) urine samples were collected from 595 patients with urothelial carcinoma. All samples were profiled by WGS. All tissue samples were pathology-checked. ecDNA and other fSCNAs were reconstructed using AmpliconArchitect software, utilizing biopsy-WGS data with CN calls as input. **B**, Fraction of patients bearing ecDNA in four different sample types. **C**, Proportion of tumor samples with ecDNA, stratified by ecDNA status in AT samples. The *P* value was calculated using a one-sided Fisher exact test. OR and 95% CIs were also determined. **D**, Number of identical ecDNA sequences detected in both ATs and their corresponding paired tumor samples. **E**, Comparison of the maximum CN segments from six ecDNAs detected in both ATs and their corresponding paired tumor samples. The *P* value was calculated using a paired-sample *t* test. **F** and **G**, AmpliconArchitect-generated SV and breakpoint graph (**F**) and the circular structure (**G**) of the *ERBB2*-containing ecDNA identified in urothelial dysplasia and tumor samples from patient CCGA-032. Similarity score and *P* value were computed based on the genomically overlapping regions of two ecDNA amplicons (see "Methods"). **H**, Representative hematoxylin and eosin-stained images of AT (pathology-confirmed urothelial dysplasia) and tumor samples from patient CCGA-032. Signiliarity score and *P* value were of the targeting *ERBB2* gene loci (**I**) and Circle-seq (**J**). The centromere of chromosome 17 (CEN 17) probe is represented in green, the *ERBB2* probe is highlighted in red, and the nuclei counterstained with 4/6-diamidino-2-phenylindole (DAPI, blue); Scale bars, 10 µm. (**A**, Created with BioRender.com.)

ecDNA⁺ patients harbored two or more distinct ecDNAs within the same tumor (Supplementary Fig. S2E). Of note, our analysis might underestimate the true ecDNA diversity within the samples because we restricted consideration to regions with high CNs in bulk sequencing and cannot fully discern genomically overlapping, but structurally distinct ecDNA species.

We then examined the genes frequently encoded on ecDNA and found that oncogenes are important components of ecDNA in urothelial carcinoma. Specifically, we observed that the ecDNA contained a range of known oncogenes, including CCND1, FGF4, FGF3, and MDM2, which have well-established roles in urothelial carcinoma development (Supplementary Fig. S3A and SB; Supplementary Tables S13 and S14). Moreover, we found a significant enrichment of oncogenes within ecDNA compared with non-oncogenes (Supplementary Fig. S3C; two-sided Fisher exact test, P = 1.23e-05), and 35% of the ecDNAs contained multiple oncogenes on the same molecule (Supplementary Fig. S3D). We also found that in patients with only one ecDNA, the proportion of ecDNAs carrying oncogenes is significantly higher than in patients with multiple ecDNAs (Supplementary Fig. S3E; two-sided Fisher exact test, P = 0.001). On average, ecDNAs harbored 1.7 unique oncogenes per amplicon (740 oncogenes in 440 ecDNAs) compared with 0.7 unique oncogenes per amplicon in other fSCNAs (1,391 oncogenes in 2,028 other fSCNAs). After adjusting for amplicon size, ecDNA carries an average of 1.1 oncogenes per megabase, whereas other fSCNAs carry only 0.9 oncogenes per megabase. Collectively, these data demonstrate a positive selection for oncogene-containing ecDNA in urothelial carcinoma.

Importantly, the CN and RNA expression levels of oncogenes were significantly higher in tumors in which oncogenes were amplified by ecDNA compared with samples in which they were amplified on other fSCNAs (Supplementary Fig. S3F; Wilcoxon rank-sum test; CN, P = 4.93e-78; mRNA expression, P = 0.0004). However, after normalizing gene expression levels by gene CN, no significant difference was observed in the expression of genes encoded by ecDNA compared with those encoded by other fSCNAs (Supplementary Fig. S3F; Wilcoxon rank-sum test; P = 0.5825). These findings support the hypothesis that ecDNA can drive tumorigenesis by promoting substantial oncogene expression, with CNA likely the primary underlying mechanism (Supplementary Fig. S3G; refs. 12, 38, 39).

ecDNA Can Arise in Flat Urothelial Lesions

Previous studies have shown that macroscopic somatic clonal expansion can occur within morphologically "normal" urothelium in patients with urothelial carcinoma (29). Moreover, adjacent tumor tissues have been characterized as a distinct intermediate state between healthy and malignant tissue (40). To understand whether ecDNA may be linked to early urothelial oncogenesis, we examined tissue biopsies taken from areas adjacent to the frank, raised tumor tissue, termed ATs. We identified 12 distinct ecDNA species in nine AT samples (Fig. 1B; Supplementary Table S9). Among the nine ecDNA⁺ AT samples, six exhibited histologic features that could potentially be precursor lesions for invasive

cancer (1, 2), including hyperplasia (one case), dysplasia (two cases), and carcinoma in situ (three cases; Supplementary Fig. S4A). Phylogenetic analysis revealed a common origin for all ecDNA⁺ AT samples and their paired tumors within the same patients (Supplementary Fig. S4B). The presence of ecDNA in tumors was significantly elevated when paired with ecDNA⁺ ATs compared with ecDNA⁻ ATs (Fig. 1C; onesided Fisher exact test, P = 0.037). To compare the ecDNA structure among ATs and paired tumors, we calculated the similarity score between amplicons by evaluating the relative proportions of shared overlap in genomic coordinates and structural variation breakpoint locations, as previously described (Supplementary Table S10; ref. 30). Among these 12 ecDNAs in ATs, six had similar genic content as ecDNA detected in corresponding paired tumors (Similarity-P < 0.05), whereas the remaining six did not persist in the tumor samples (Fig. 1D). Although the sample size was limited, paired comparisons revealed a significantly higher ecDNA CN in tumors than in ATs (Fig. 1E; paired-sample t test, P = 0.028). For instance, in the case of CCGA-032, we detected an ecDNA containing several oncogenes, such as CDK12, ERBB2, and RARA, in both AT and tumor samples (pT2a, high grade; Fig. 1F–H; similarity score, 0.875; P = 5.06e-07). The maximum CN of ecDNA segments, estimated from WGS data in AT and tumor samples, was 7 and 80, respectively. To further validate the presence of ecDNA and its CN, we performed DNA FISH in interphase nuclei on a formalin-fixed, paraffin-embedded (FFPE) section targeting the ERBB2 gene loci and analyzed the corresponding Circle-seq (a circular DNA enrichment library) data. As expected, we observed a robust fluorescence signal, with a pattern of ERBB2 foci consistent with ecDNA, and a high Circle-seq read density at ERBB2-ecDNA regions in both AT and tumor samples (Fig. 1I and J). Although limited in ecDNA⁺ AT sample size, these data suggest that ecDNA can arise in flat urothelial lesions and be selected during the malignant transformation and progression of urothelial carcinoma, similar to the transition from precancer to cancer in patients with Barrett's esophagus (30).

ecDNA Contributes to Clonal Evolution through Structural Rearrangements during Urothelial Carcinoma Malignant Transformation and Recurrence

Cancers evolve through a reiterative process of clonal expansion, genetic diversification, and clonal selection within the adaptive landscape of the tumor microenvironment (41). To investigate whether ecDNA contributes to the clonal evolution of urothelial carcinoma by undergoing diversifying mutations and subsequent selection, we analyzed the structure and oscillating CN states of six pairs of ecDNAs in which the same ecDNA was detected in both AT and tumor samples from the same patients (Supplementary Fig. S5A). We observed that ecDNA confers tremendous temporal intratumoral heterogeneity. For instance, in the case of CCGA-075, an ecDNA (ecDNA-1) containing MDM2 and YEATS4 was identified in the AT sample. This ecDNA was also present in the paired tumor sample but with reduced structural complexity and fewer oscillating CN states (Fig. 2A; Supplementary Fig. S5A and S5B;



D

ecDNAs identified in patients CUGA-058, CUGA-077, and CUGA-121

Patient	Primary tumor (P)	Recurrent tumor (R)	Similarity score (P value)	Complexity score (P/R)	No. of CN states (P/R)
CUGA-058	T_amplicon15_ecDNA_1	T-Rp_amplicon5_ecDNA_1	0.5225 (0.0060)	4.9494/0.6847	4/1
CUGA-058	T_amplicon13_ecDNA_1	/	/	5.0843/none	3/none
CUGA-058	T_amplicon6_ecDNA_1	/	/	5.7637/none	26/none
CUGA-058	/	Rp_amplicon1_ecDNA_2	/	none/2.8632	none/2
CUGA-077	T_amplicon2_ecDNA_1	Rp2_amplicon1_ecDNA_3	0.4063 (0.0028)	3.9113/4.9476	5/4
CUGA-077	/	Rp2_amplicon1_ecDNA_1	/	none/5.5574	none/10
CUGA-077	/	Rp2_amplicon1_ecDNA_2	/	none/4.1419	none/9
CUGA-077	/	Rp2_amplicon2_ecDNA_1	/	none/4.1335	none/2
CUGA-121	/	Rp_amplicon2_ecDNA_1	/	none/3.6415	none/3
CUGA-121	/	Rp_amplicon2_ecDNA_2	/	none/4.5932	none/7
CUGA-121	/	Rp_amplicon2_ecDNA_3	/	none/1.3662	none/1

Figure 2. ecDNA evolves through structural rearrangements during urothelial carcinoma malignant transformation and recurrence. **A**, Analysis of ecDNA evolution. SV and breakpoint graph generated by AmpliconArchitect for ecDNA detected in patient CCGA-075. We identified an ecDNA containing *MDM2* and *YEATS4* in ATs. This ecDNA was also present in the paired tumor sample but exhibited lower structural complexity and a lower number of oscillating CN states. Moreover, in tumor tissue, we observed a second ecDNA (ecDNA-2) containing *HRAS* and *CCND1* within the same amplicon. **B**, A timeline of the clinical history of the four patients with matched primary and recurrent tumors. Representative CT and hematoxylin and eosin-stained images are shown. The tumor grade, stage, and ecDNA status are annotated alongside the hematoxylin and eosin image. TURBT, transurethral resection of the bladder tumor; BCG, Bacillus Calmette-Guérin; RC, radical cystectomy. **C**, Phylogenetic trees were constructed from all somatic nonsynonymous mutations. The branch lengths are proportional to the number of mutations. **D**, Complexity score and the number of oscillating CN states of ecDNAs detected in patients CUGA-058, CUGA-077, and CUGA-121.

complexity score: AT-ecDNA-1, 5.4; tumor-ecDNA-1, 2.5; CN state: AT-ecDNA-1, 8; tumor-ecDNA-1, 3). Moreover, tumor samples exhibited a second ecDNA (ecDNA-2) containing

HRAS and *CCND1* within the same amplicon. This indicates that ecDNA can evolve through structural rearrangement during tumor development.

In our cohort, three patients (CUGA-058, CUGA-077, and CUGA-121) whose tumor samples carried ecDNA had matched primary and recurrent tumor samples with available WGS data (Fig. 2B). Tumor evolutionary tree analysis reveals that paired primary and recurrent tumors share some early somatic mutations, indicating that recurrent tumors are not entirely "new" tumors (Fig. 2C). In one case (CUGA-121), ecDNA was absent in the primary tumor sample but present in the recurrent tumor sample (Fig. 2D). In the second case (CUGA-058), the primary tumor samples had three ecDNAs. However, most ecDNA segments were not detected in the recurrent tumor samples. Instead, a novel ecDNA was identified in the recurrent tumor samples (Fig. 2D; Supplementary Fig. S5C). In the third case (CUGA-077), the primary biopsy showed an ecDNA amplification of MDM2. The ecDNA of the relapsed tumor contained most regions from the ecDNA of the primary tumor but exhibited higher structural complexity (Supplementary Fig. S5C). Additionally, three novel ecDNAs were identified in the recurrent tumor (Fig. 2D; Supplementary Fig. S5C). Although limited to three cases, these observations provide additional evidence that ecDNA experiences structural changes during tumor development and that new ecDNAs may emerge during the process.

Clonal and Subclonal Evolution of ecDNA in Urothelial Carcinoma Development

The presence of ecDNA in AT samples prompted us to further explore the evolutionary dynamics of ecDNA in urothelial carcinoma. We analyzed the WGS data from 211 spatially distinct tumor regions and 52 AT/blood samples obtained from 57 patients with urothelial carcinoma (with two to six tumor regions per patient; five patients lacked an available germline sample; Fig. 3A; Supplementary Figs. S6 and S7; Supplementary Table S15). In total, we identified 59 unique ecDNA species across 26 patients (Fig. 3A and B). In contrast to most urothelial carcinoma drivers, we found that ecDNA displayed pronounced intratumor heterogeneity (Supplementary Fig. S8A). In 11 cases, ecDNA was detected only in specific tumor regions rather than uniformly across all regions (Fig. 3A). We identified a total of 15 unique clonal ecDNAs (the same ecDNA lineage present in all tumor regions) in 13 of 57 patients (Fig. 3B). For instance, in CUGA-MR-018 (a patient with multifocal bladder tumors), in whom a chimeric ecDNA from chromosomes 6 and 11 with CCND1 was shared across every tumor region (Supplementary Fig. S6A-S6D). These data establish a direct link between ecDNA and the evolution of multifocal cancer. However, the majority of ecDNAs (75%, 44 of 59) were identified as subclonal ecDNA (present in at least one, but not all, regions), and subclonal ecDNA was observed in 19 of 57 patients (Fig. 3B). For example, in CUGA-MR-027, a heterogeneous ecDNA containing CCND1 was present only in tumor regions R2 to R4 and R6 but not in R5. However, chromosomal CCND1 amplifications through complex noncyclic amplification were detected in tumor region R5, which exhibited a lower CN and structural complexity in the fSCNA region (Supplementary Fig. S7A-S7C). Additionally, constructing a phylogenetic tree based on all somatic nonsynonymous mutations of all tumor regions indicates that R5 represents an earlier stage of tumor development

(Supplementary Fig. S7D). These findings suggest that multiple topologies of focal amplifications emerge during tumor progression, providing a vivid example of the dynamic evolution and diversification of focal amplifications.

Six tumors had both clonal and subclonal ecDNA events (Fig. 3B). For example, in CUGA-MR-032, a clonal ecDNA (ecDNA-1) was found in all tumor regions (R1-R4), whereas two subclonal ecDNAs (ecDNA-2 and ecDNA-3) were present only in tumor regions R3 and R4 (Fig. 3C and D). Clonal and subclonal ecDNAs collectively shaped the urothelial carcinoma evolution (Fig. 3E). We then assessed intratumoral heterogeneity at the patient level by calculating the proportion of genes exhibiting copy-number variations (CNV) across samples, relative to all genes. Our findings indicate that patients with subclonal ecDNA demonstrate higher levels of intratumoral heterogeneity compared with those without ecDNA or those harboring only clonal ecDNA (Fig. 3F; Supplementary Table S15). However, the difference between patients with only subclonal ecDNA and those with only clonal ecDNA was not statistically significant, likely due to the limited sample size (Fig. 3F; Wilcoxon rank-sum test, P = 0.058). Moreover, we observed that almost all clonal ecDNAs (14 of 15) carried oncogenes, whereas nearly 43% (19 of 44) of subclonal ecDNAs were devoid of canonical oncogenes (Fig. 3G; two-sided Fisher exact test, P = 0.01124). The ratio of kataegic mutation clusters carried on clonal ecDNA and subclonal ecDNA did not show a significant difference (Supplementary Fig. S8B; twosided Fisher exact test, P = 0.4725).

Mutation mapping on detected ecDNAs enabled us to infer the relative timing of mutations in the context of clonal and subclonal ecDNA evolution (9, 19). By examining variant allele frequencies (VAF), we found that mutations on clonal ecDNAs exhibited significantly higher VAFs than those on subclonal ecDNAs (Supplementary Fig. S8C; Wilcoxon rank-sum test, P = 0.0055), suggesting that clonal ecDNAs underwent longer periods of selection. Furthermore, our analysis revealed that most driver gene mutations had a higher VAF than both clonal and subclonal ecDNAs (Supplementary Fig. S8D), suggesting they occurred early in tumor progression.

Overall, these data suggest that ecDNA formation may be an ongoing process, providing a substrate for subclonal competition and selection, including for oncogenes, during tumor progression (Fig. 3H).

ecDNA Is Associated with Shortened Patient Survival and an Unstable Cancer Genome

After revealing the properties and evolutionary trajectory of ecDNA in urothelial carcinoma, a crucial next step is understanding how its presence correlates with clinical manifestations. We utilized samples from 488 patients with paired tumor–germline WGS data to minimize potential biases for subsequent analysis. We observed a significant variation in the ecDNA frequency across different urothelial carcinoma subtypes. Specifically, ecDNA⁺ patients more frequently exhibited a high-grade and muscle-invasive (pT2-4) histologic subtype (two-sided Fisher exact test; P < 0.01). Renal pelvis tumors demonstrated a lower ecDNA frequency than bladder tumors (P = 0.015). No significant differences in ecDNA



Figure 3. Spatial dissection of ecDNA in urothelial carcinoma. **A**, Overview of the tumor multiregional WGS cohort (*n* = 211 tumors from 57 patients are shown). Each column represents one tumor region, displaying patient ID, tumor location, and genetic alterations. Genetic alterations included ecDNA, breakage-fusion-bridge (BFB), linear amplification (AMP), complex AMP, chromothripsis, whole-genome duplication (WGD), as well as somatic mutation of drivers. Potential drivers of urothelial carcinoma were inferred from WGS data using three different methods: 20/20+, MutSigCV, and dNdScv (see "Methods"). INDEL, insertions and deletion; LNM, lymph node metastasis. **B**, The top illustrates the proportion of clonal ecDNA. The bottom displays the fraction of patients with indicated ecDNA events. Clonal ecDNA was defined as the same ecDNA detected in all tumor regions from the same patient. Subclonal ecDNA was defined as the ecDNA present in at least one, but not all, regions from the same patient. Most ecDNAs are present in the subclonal form. **C**, Representative images of the multiregion sampling site of patient CCGA-MR-032. ecDNA detected in every tumor region was also annotated. **D**, The circular structure of one clonal ecDNA (ecDNA-1) and two subclonal ecDNAs (ecDNA-2 and -3) identified in tumors from patient CCGA-MR-032. **E**, Phylogenetic trees of patient CCGA-MR-032 were constructed from all somatic nonsynonymous mutations. The branch lengths are proportional to the number of mutations. Potential driver genes and ecDNAs are labeled. **F**, Comparison of intratumoral heterogeneity index was calculated as the proportion of (continued on following page)

frequency were observed based on age (P = 0.452), sex (P = 0.750), or smoking history (P = 0.695); Fig. 4A; Supplementary Fig. S9A). These results imply that specific physiologic cues and the tumor microenvironment can repress or fuel ecDNA-driven carcinogenesis. Patients with tumors harboring ecDNA, regardless of the specific gene amplified, experienced significantly worse outcomes than those without any fSCNAs (Fig. 4B; log-rank test, P = 3.0E-05). Patients with tumors harboring other fSCNAs had similar survival impacts (Fig. 4B; P = 0.007). A multivariate analysis that additionally considered tumor stage showed that the presence of ecDNA was significantly associated with an increased hazard ratio (Fig. 4C; P = 0.032, ecDNA vs. non-fSCNA; P values determined by multivariate Cox proportional hazards model). The ecDNA structure and whether it carries oncogenes were not associated with overall survival in urothelial carcinoma (Supplementary Fig. S9B; P > 0.1).

Given the substantial adverse impact of ecDNA on patient outcomes, we next investigated the genetic features associated with ecDNA (Supplementary Fig. S9C-S9F). After adjusting for tumor stage, we found that ecDNA⁺ tumors exhibited an overall increase in structural variants (SV) and higher chromosomal instability, as quantified by the ploidy, the fraction of genome altered, and the proportion of wholegenome duplication and chromothripsis events, compared with tumors without focal amplifications and tumors with other fSCNAs (Fig. 4D). Moreover, compared with tumors without fSCNA, ecDNA+ tumors showed a higher tumor mutation burden and number of predicted neoantigens (Fig. 4D; Supplementary Fig. S9D). Because not all somatic mutations represent potential therapeutic targets, we conducted detailed annotations using OncoKB (a precision oncology knowledge base; ref. 42) for all detected mutations. We observed a lower frequency of level 1 actionable alterations in ecDNA+ tumors (Fig. 4D; Supplementary Fig. S9E), suggesting ecDNA⁺ patients have a diminished potential for benefiting from FDA-approved drugs. Apolipoprotein B mRNA-editing enzyme catalytic polypeptide 3 (APOBEC3) may contribute to ecDNA evolution (36). To test this, we determined mutational signatures in the urothelial carcinoma cohort and found a higher prevalence of APOBEC-related signatures of SBS2 and SBS13 in ecDNA⁺ tumors (Supplementary Fig. S9F; Wilcoxon rank-sum test, P < 0.05).

Inactivating mutations in *TP53* are known to promote genome instability (43-45). Consistent with this, after adjusting for tumor stage, we observed a higher frequency of *TP53* alterations in ecDNA⁺ tumors compared with tumors without fSCNA (Fig. 4E), implying that p53 malfunction may contribute to ecDNA formation in urothelial carcinoma. In contrast, some well-established favorable molecular prognostic indicators, including *FGFR3*, *HRAS*, and *STAG2*, were less frequently altered in ecDNA⁺ tumors (Fig. 4E; Supplementary Table S16). *TP53* mutatations and *MDM2/4* amplification were mutually exclusive (Fig. 4F; $\chi^2 = 0.001637$). Moreover, the majority (86%) of ecDNA⁺ tumors carry alterations in the *TP53* pathway (Supplementary Fig. S9C).

Collectively, these data suggest that ecDNA may play a role in urothelial carcinoma progression and is associated with genomic instability and aggressive biological features, consistent with recent pan-cancer ecDNA findings (19).

ecDNA Is Associated with a Distinct TIME

The TIME plays a crucial role in urothelial carcinoma development, with increasing immune infiltration as the disease advances (46). Previous pan-cancer studies have explored the relationship between ecDNA and TIME using bulk RNA-seq data (47, 48); however, bulk RNA-seq has limitations in dissecting the interplay between the host immune system and ecDNA-containing cancer cells. To directly elucidate the differences in the immune ecosystem between ecDNAand ecDNA⁺ tumors, we analyzed scRNA-seq data from 19 tumors (10 ecDNA⁻, including 2 with other fSCNA and 8 without any fSCNA, and 9 ecDNA+; Fig. 5A). After stringent filtering (see "Methods"), we obtained 259,808 single-cell transcriptomes (Supplementary Table S17). Based on lineagespecific marker genes, cells were classified into six major clusters (Fig. 5A; Supplementary Fig. S10A). Single-cell CNV analysis using inferCNV software (49) identified all epithelial cells as malignant (Supplementary Fig. S10B). We observed that ecDNA⁺ tumors had a lower proportion of malignant cells and a greater enrichment of immune cells compared with ecDNA- tumors (Fig. 5B; Supplementary Fig. S10C and S10D). To avoid potential technical artifacts, we performed IHC staining (n = 19 tumors) and multiplexed immunofluorescence (mIF) staining analysis (Supplementary Fig. S10E-S10G; Supplementary Table S18). The immune cell proportions determined by scRNA-seq were positively correlated with the immune infiltration patterns observed in IHC staining (Supplementary Fig. S10F; Pearson's r = 0.575, P = 0.0113). To further validate this observation, we used the CIBERSORTx method to estimate the proportion of immune cells in tumors from bulk RNA-seq data (n = 331tumors). We revealed a significant increase in the density of immune cells within tumors harboring ecDNA compared with those without fSCNA (Fig. 5C; Wilcoxon rank-sum test, P = 0.00029). No significant difference in immune cell infiltration was observed between tumors with ecDNA and those with other fSCNAs (Fig. 5C; P = 0.65). This significant difference in immune cell infiltration between ecDNA⁺ tumors and fSCNA- tumors persisted after adjusting for tumor stage (Fig. 5D). These data suggested that ecDNA⁺ tumors harbor a distinct tumor immune ecosystem.

Figure 3. (Continued) genes with CNVs between samples, out of all genes (see "Methods"). The P value was calculated using a Wilcoxon rank-sum test. G, Fraction of ecDNAs carrying oncogenes, stratified by ecDNA clonal status. The P value was calculated using a two-sided Fisher exact test. H, A schematic diagram showing the evolutionary trajectory of ecDNA during the cancer life history. Somatic variants were classified as initiating clonal variants, late clonal variants, and late subclonal variants according to subclonal diversification. The normal urothelium acquires somatic mutations in driver genes when stimulated by carcinogens. Subsequently, some ecDNAs may appear in the early stages of tumors, even in precancerous lesions, and persist throughout cancer progression. However, the majority of ecDNAs emerge during the advanced stages of the tumor, contributing to subclonal evolution. Clonal and subclonal ecDNA collectively shaped the urothelial carcinoma evolution.



Figure 4. Clinical and genomic features associated with ecDNA. **A**, Oncoprint illustrating clinical and genomic attributes between the 488 tumors categorized into three groups based on fSCNA status: without fSCNA (*n* = 210 patients), without non-ecDNA-based amplifications (other fSCNA, *n* = 93 patients), and with ecDNA amplification (*n* = 185 patients). Each column represents a patient, ordered by the number of SVs from highest to lowest. For patients with multiregional WGS data, we calculated the average value of each parameter across all tumors for statistical analysis. AMP, amplification; BFB, breakage-fusion-bridge; FGA, the fraction of genome altered; TMB, tumor mutation burden; WGD, whole genome duplication. **B**, Kaplan-Meier plot depicting the overall survival of 488 patients with urothelial carcinoma stratified by fSCNA status. The *P* value was calculated using a log-rank test. **C**, Forest plot depicting the associations between the presence or absence of ecDNA and various genomic characteristics, such as the number of SVs, ploidy status, FGA, WGD events, chromothripsis, TMB, neoantigen load, and the frequency of OncoKB level 1 mutation. The analysis was adjusted for the tumor stage. Error bars represent 95% Cls for the OR estimates. OncoKB level 1 mutation: "FDA-recognized" biomarker predictive of response to an "FDA-approved drug" in this indication. **E**, Forest plot illustrating the results of a regression model that examines the association between the presence of ecDNA amplifications in tumors from the entire urothelial carcinoma cohort (*n* = 488 patients). The analysis was adjusted for the tumor stage. Error bars represent 95% Cls for the OR estimates of a regression model that examines the association between the presence of ecDNA amplifications and the odds of harboring high-impact driver gene mutations in tumors from the entire urothelial carcinoma cohort (*n* = 488 patients). The analysis was adjusted for the tumor stage. Error bars represent 95% Cls for the OR estimates of a regression

The reclustering of immune cell populations identified multiple subsets of myeloid cells, B cells, and T/NK cells (Fig. 5E; Supplementary Fig. S10H). We observed significant variation in the distribution of these clusters across fSCNA subgroups (Fig. 5F). Specifically, ecDNA⁺ tumors displayed enrichment of regulatory T cells (Treg), whereas ecDNA⁻ tumors harbored more macrophages (Supplementary Fig. S10I). Tregs are known to enhance the immunosuppressive program of CD4⁺ T cells, promoting tumor development (50). Moreover, we found a significantly higher prevalence of

RESEARCH ARTICLE



Figure 5. TIME associated with ecDNA. A, Uniform Manifold Approximation and Projection (UMAP) plots of 259,808 cells profiled by scRNA-seq colored by major cell types (left) and fSCNA status (right). B, Bar plot depicting the relative proportions of each identified major cell type across groups. **C**, Comparison of immune cell fractions across three groups of tumors: fSCNA-negative (n = 142), other fSCNA-positive (n = 64), and ecDNA-positive (n = 125). The abundance of immune cell types was estimated from bulk RNA-seq data using CIBERSORTx, an established deconvolution method. The P value was calculated using a Wilcoxon rank-sum test. D, Forest plot illustrating ORs for increased immune cell infiltration associated with the presence of ecDNA or other fSCNA, adjusted for tumor stage. Error bars represent 95% CIs for the OR estimates. *, P < 0.05; **, P < 0.01. E, UMAP plot showing the major immune cell type in the urothelial carcinoma ecosystems. F, UMAP visualization of major immune cell lineages (top) and cell density (bottom) across three groups. High relative cell density is shown as bright magma. G, Expression levels of functional markers across B-cell clusters. H, UMAP plot of 158,450 malignant cells colored by patient ID (n = 13). Tumor samples (n = 6) containing fewer than 20 malignant cells were excluded from the analysis. I, Bar plot showing the enrichment of specific pathways, based on CancerSEA and HALLMARK gene sets of DEGs of malignant cells between ecDNA- (including both of fSCNA- and other fSCNA*) and ecDNA+ tumors. J, Violin plot showing the APM score across three groups of tumors. The P value was calculated using a Wilcoxon rank-sum test. K, Fraction of tumors carrying allele-specific HLA-I LOH alterations across three groups of tumors. HLA-I LOH were inferred from WGS data using the SpecHLA software. The P value was calculated using a two-sided Fisher exact test. L, Forest plot illustrating the OR for increased frequency of HLA-I LOH events associated with the presence of ecDNA, adjusted for tumor stage. Error bars represent 95% Cls for the OR estimates. *, P < 0.05. M, Bubble chart showing the MHC-I and CD8 ligand-receptor pairs between malignant cells and three T-cell types across three groups of tumors. The color gradient represents the interaction strength. N, Diagram showing how the ecDNA-containing malignant cells could evade immune cell attacks by downregulating the expression of MHC-I. Commun. Prob., communication probability; MUT, mutated; WT, wild-type. (Created with BioRender.com.)

B-cell infiltration in ecDNA⁺ tumors compared with ecDNA⁻ tumors. We identified a total of 15 distinct B-cell clusters, including one for cycling plasma cells, 11 for plasma cells, one for naïve B cells, and three for memory B cells (Supplementary Fig. S10I). Of note, memory B cells expressed higher levels of several regulatory B cell-related molecules, including *IL10*, *TGFB1*, *CD40*, *CD24*, and *CD27*, further reflecting the immunosuppressive nature of the ecDNA-associated urothelial carcinoma TIME (Fig. 5G).

During tumorigenesis, cancer cells evolve to avoid or suppress immune attacks through various mechanisms (51). The distinct clinical and molecular features associated with ecDNA prompted us to hypothesize that these distinctions arise from intrinsic variations within the malignant cells themselves. To explore this hypothesis, we analyzed scRNA-seq data from 158,450 malignant cells across 13 tumors. Six tumors with fewer than 20 malignant cells were excluded from the analysis (Supplementary Table S17). These malignant cells exhibited significant intertumor heterogeneity and were classified into 13 distinct clusters based on their tumor origin (Fig. 5H). After adjusting for patient-specific effects, we observed a distinct distribution of malignant cells across fSCNA groups (Supplementary Fig. S11A). Using nonnegative matrix factorization (NMF) on each malignant cell, we identified five distinct expression programs that reflect common patterns of intratumor heterogeneity across different tumors (Supplementary Fig. S11B; Supplementary Table S19). Each expression program represents a unique biological process and varies among malignant cells from each group (Supplementary Fig. S11C; Supplementary Tables S20 and S21). Differential gene expression analysis between ecDNA- and ecDNA+ malignant cells revealed an enrichment of genes associated with the cell cycle in ecDNA+ tumors, whereas genes linked to the p53 pathway were predominant in ecDNA- tumors (Fig. 5I; Supplementary Table S22). Furthermore, ecDNA+ tumors exhibited significantly higher CNV scores compared with ecDNA- tumors (Supplementary Fig. S11D; Wilcoxon rank-sum test; P < 2.2e-16). These findings are consistent with our genomic analysis and further support the presence of distinct transcriptional programs in ecDNA⁺ malignant cells.

Besides abundance, the spatial distribution of immune cells-especially CD8⁺ T cells-is closely linked to TIME stratification. Lower intraepithelial CD8+ T-cell density is associated with poorer prognosis in urothelial carcinoma (46). Our analysis of genes involved in the antigen presentation machinery (APM) revealed the downregulation of MHC class I (MHC-I) genes and B2M expression in malignant cells in ecDNA⁺ tumors relative to tumors without fSCNA (Fig. 5J; Wilcoxon rank-sum test; APM score, P < 2.2e-16). Furthermore, by analyzing WGS data with SpecHLA software (52), we observed that allele-specific loss of heterozygosity (LOH) of class I HLA genes (HLA-I LOH) occurred more frequently in ecDNA⁺ tumors than in tumors without fSCNA (Fig. 5K). Controlling for tumor stage, ecDNA was strongly associated with HLA-I LOH (Fig. 5L; OR 2.1, 95% CI, 1.3-3.7). Ligandreceptor analysis indicated that the cross-talk between malignant cells and CD8+ T cells, IFN+ T cells, and CD56 $^{\rm dim}{\rm CD16^{\rm hi}}$ NK cells through the MHC-I:CD8 axis was diminished or potentially absent in ecDNA+ tumors (Fig. 5M). To further elucidate the spatial interplay between CD8+ T cells and malignant cells, we performed mIF on four tumor samples. We observed that tumors with high MHC-I expression exhibited either an "infiltrated" or "desert" immune phenotype, characterized by the infiltration of CD8+ T cells into the tumor urothelium or the absence of CD8⁺ T cells within the TIME, respectively. In contrast, tumors with low MHC-I expression exhibited an "excluded" immune phenotype, with CD8+ T cells predominantly localized in the surrounding stroma, or a "desert" immune phenotype (Supplementary Fig. S11E). Further DNA FISH experiments confirmed the presence of ecDNA in two selected tumor samples (Supplementary Fig. S11F). These data suggest that one possible mechanism by which ecDNA⁺ tumors evade immune cell attacks is through the reduction of MHC-I antigen presentation, leading to dysfunctional antitumor immunity in ecDNA-driven urothelial carcinoma (Fig. 5N). It would be of interest to explore immunotherapeutic approaches for reinvigorating the immune system against ecDNA⁺ tumors.

Urinary Sediment-Derived DNA Has High Specificity in Detecting ecDNA

Urine serves as an ideal clinical source for liquid biopsy (10, 53). However, the full potential of urine in identifying tumor-specific ecDNA remains to be fully understood. We analyzed the WGS data of paired tumor and urinary sediment samples from 167 patients with urothelial carcinoma. Despite lower tumor purity in urinary sediment, resulting in a reduced ecDNA detection rate (23%, 39 of 167) compared with tumors (46%, 77 of 167; Fig. 6A; tumor purity, Paired t test, P = 3.79e-23; ecDNA frequency, Fisher exact test, P = 9.47e-05), sequencing of urinary sediment-derived DNA demonstrated robust specificity (95.6%) in ecDNA detection (Fig. 6B). This high specificity remained consistent across different clinical subgroups and varying tumor purities in urinary sediment (Fig. 6C). Moreover, sequencing of urinary sediment-derived DNA revealed a moderate sensitivity (45.5%) for ecDNA detection, primarily influenced by tumor purity within the urinary sediment and tumor location (Fig. 6B and C). Overall, these data indicate that urinary sediment-derived DNA sequencing offers a promising approach for ecDNA detection, with a positive predictive value and negative predictive value of 89.7% and 67.2%, respectively (Fig. 6B).

At the ecDNA level, we identified 65 distinct ecDNAs in urine samples, exhibiting significant genomic overlap with tumor-derived ecDNA (n = 181; Fig. 6D; ISTAT test, P = 3.3184e-09). Of these, 62% (40 of 65) of ecDNAs were shared between paired tumor-urine samples with high similarity scores (similarity-P < 0.05), whereas 25 ecDNAs were exclusively detected in urine (Fig. 6E and F). Notably, among the 25 urine-specific ecDNAs, five were identified in paired tumor tissues as non-ecDNA-based amplifications, four overlapped with tumor ecDNAs, and two overlapped with non-ecDNA-based amplifications (Fig. 6G; Supplementary Fig. S12A and S12B). Furthermore, 14 urinary ecDNAs showed no overlap with any fSCNA in tumors. These ecDNAs exhibited shorter lengths and a decreased likelihood of



Figure 6. Urinary sediment-derived DNA has high specificity in detecting ecDNA. **A**, Tumor purity and focal amplification profiles of paired tumor and urine samples (*n* = 167). Each column represents one patient. Each row represents somatic variants or focal amplification detected in preoperative urinary sediment (indicated by "U") and paired tumors (indicated by "T"). The *P* value was calculated using the Wilcoxon rank-sum test and paired-sample t test. **B**, Confusion matrix of the urinary sediment-derived DNA for ecDNA detection in 167 patients with urothelial carcinoma. **C**, Sensitivity and specificity of urinary sediment-derived DNA in detecting ecDNA, stratified by clinical features and tumor purity within the urinary sediment. **D**, Genomic distribution of ecDNA detected in urine and tumor samples. The ecDNA intervals were identified from AmpliconClassifier. The overlap's statistical significance between urine-ecDNA and tumor-ecDNA was calculated using an iSTAT test (https://github.com/shahab-sarmashghi/ISTAT). **E**, Classification of urinary ecDNA. (*n* = 65 ecDNAs). T⁺ U⁺: ecDNA detected in both tumor and urine samples (*n* = 40 ecDNAs). T⁻ U⁺: ecDNA detected sclusively in urine (*n* = 25 ecDNAs). **F**, Similarity score of 40 paired ecDNA detected in urine and matched tumor samples. Similarity scores and *P* values were calculated based on the genomic overlap between the two ecDNA amplification; BFB, breakage-fusion-bridge; WGD, whole-genome duplication.

harboring oncogenes compared with ecDNAs identified in both urine and tumor samples (Supplementary Fig. S12C and S12D). These findings suggest the presence of minor ecDNA clones in urine and also further emphasize that multiple forms of focal amplifications can emerge during tumor progression.

DISCUSSION

The human genome consists of 23 pairs of linear chromosomes. In cancer, cancer-causing genes can liberate themselves from chromosomes and relocate to ecDNA, thereby reshaping the genetic intratumoral heterogeneity and promoting tumorigenesis (11, 19, 20, 22, 26, 54). Recently, Nguyen and colleagues (9) demonstrated that the interplay between mutagenesis and ecDNA may play a role in the evolution of treatment resistance in urothelial carcinoma. However, the evolutionary trajectory of ecDNA and whether the evolution of ecDNA itself can contribute to the early clonal expansion in cancer remain largely unexplored. Meanwhile, the spatial dynamics of ecDNA and the impact on genetic heterogeneity is still poorly understood. In this study, by analyzing the largest urothelial carcinoma WGS dataset to date, we demonstrate that ecDNA can be an early clonal mutational event in the transformation from flat urothelial lesions to invasive cancer. Throughout this process, ecDNA undergoes complex structural changes, supporting a hypothesis that tumors evolve through the continuous selection of such ecDNA. Although we cannot exclude the possibility that individual cancer cells have invaded from the adjacent tumor, these findings expand our understanding of the diversity of genomic alterations during the malignant transformation process of urothelial carcinoma and suggest the potential for early disease intervention in patients bearing ecDNA. A larger cohort of longitudinal WGS studies from precancer to cancer is urgently needed to further understand the frequency and role of ecDNAs in precancerous tissue samples.

Traditional methods for tracking cancer clonality rely on chromosomal inheritance, making it challenging to decipher the clonal architecture and evolutionary patterns in ecDNA-driven cancers. To address this, we used a multiregion tumor sampling approach to evaluate the spatial and temporal diversity of ecDNA in the development of urothelial carcinoma. We demonstrate that ecDNA in multistage urothelial carcinoma carries different somatic variants of the same ecDNA. Deletions, translocations, and single-nucleotide variants (SNV) in one progenitor type result in the formation of new ecDNAs, which may exist either clonally or subclonally. Hence, we show that urothelial carcinoma evolves not only through incremental chromosomal mutations but also through mutations on the ecDNA, which can potentially enhance fitness at each stage of tumor progression (Fig. 3; Supplementary Figs. S5-S7). Several of the ecDNA alterations show evidence of selection, as variations over the same ecDNA recur across multiple foci within the same tumor. Also, oncogene ratios are significantly higher on ecDNA compared with other fSCNAs. Consistent with findings reported by Nguyen and colleagues (9), oncogenes such as CCND1, FGF4, and FGF3 occur recurrently on ecDNA (Supplementary Fig. S3), suggesting selection for these genes on ecDNA. ecDNA is also occasionally lost through the progression of the cancer, indicating that the loss occurs through a clonal or subclonal expansion. Loss may occur when there is no selection for the ecDNA in the local tumor microenvironment or when another advantageous mutation arises in a cancer cell that happens to lose the ecDNA through missegregation (38). Our findings also show that most ecDNAs are present in subclonal forms,

with increasing frequency in the later stages of tumor evolution. These ecDNAs provide a substrate for subclonal competition and selection, including for oncogenes, during tumor progression. Additionally, they contribute to rapid and frequent branching in the phylogenetic tree of the tumor. We also detected oncogene-containing ecDNAs in urothelial carcinoma that are under strong selection pressure. In particular, ecDNAs bearing canonical oncogenes were more likely to be clonal than oncogene-less ecDNAs. Moreover, the frequency of ecDNA increased from 27% in non-muscle-invasive disease to 46% in muscle-invasive disease, further indicating a progressive accumulation of ecDNA during tumor development.

ecDNA+ tumors generally exhibited a higher prevalence of complex karyotypes and increased chromosomal instability compared with both tumors without fSCNA and those with other fSCNAs. The majority (86%) of ecDNA⁺ tumors carry alterations in the TP53 pathway. This suggests a link between the TP53 pathway and ecDNA suppression, potentially by inhibiting their formation or maintenance. However, our data also revealed that approximately 14% of ecDNA⁺ cases did not exhibit alterations in the TP53 pathway, suggesting that ecDNA can arise in other genetic contexts. Patients for whom ecDNA was detected in at least one sample of the tumor had significantly shorter overall survival, aligning with results from previous studies of pan-cancer (11, 19, 20), small cell lung cancer (17), high-risk medulloblastoma (18), and hepatocellular carcinoma (24). Disrupting ecDNA maintenance or manipulating the formation process of ecDNA may offer promising avenues for potential therapeutic interventions (55).

Previous research suggests a direct link between ecDNA and immune evasion, as these ecDNAs may encode genes that modulate the immune response, such as suppressor of cytokine signaling 1 (SOCS1; ref. 30). Our study confirms this link by revealing striking differences in the TIME of ecDNA⁺ tumors and tumors without fSCNA using scRNA-seq data. Notably, ecDNA+ tumor cells displayed significantly lower or absent expression levels of MHC-I, and a strong association was observed between ecDNA and HLA-I LOH. This deficiency renders them invisible to the immune system, hindering their elimination. These findings highlight the potential of incorporating immune-stimulatory drugs into treatment strategies, aiming to reinvigorate the immune response against ecDNA⁺ tumors. However, significant knowledge gaps remain regarding the intricate interplay between ecDNA⁺ cancer cells and the host's immune system. A major hurdle in elucidating these mechanisms is the lack of suitable genetically engineered *in vivo* mouse models with an intact immune system. Establishing such in vivo immunocompetent will be a critical step toward a deeper mechanistic understanding of the impact of ecDNA on the immune systems and, ultimately, improved therapeutic approaches (56).

Urothelial tumors directly contact urine, providing the potential for detecting ecDNA in urine. Through analysis of paired urine and tumor samples from 167 patients with urothelial carcinoma, we found that urinary sediment-derived DNA sequencing offers a promising approach for ecDNA detection. Specifically, sequencing of urinary sedimentderived DNA demonstrated an excellent specificity (95.6%) for ecDNA detection. This high specificity was maintained

across different clinical backgrounds and varying tumor purities in urinary sediment. This is important because it means that ecDNA in urine may serve as a biomarker for urothelial carcinoma in screening, personalized treatment, and surveillance of patients. However, not all ecDNA in tumors was also recorded in urine, and sequencing of urinary sedimentderived DNA revealed a moderate sensitivity (45.5%) for ecDNA detection. The sensitivity is primarily influenced by the low volume of tumor purity within the urinary sediment, as urine samples contain not only cancer cells but also a proportion of normal cells and cellular debris (57). Thus, it may be worthwhile to consider enriching tumor cells from urine samples to enhance the sensitivity detection rate of ecDNA detection. Moreover, some of the ecDNA clones were detected only in urine samples. Because of the intratumoral genetic heterogeneity of ecDNA, and the fact that individual samples may miss specific ecDNAs, urine provides a pooled substrate for detecting the wider spectrum of ecDNAs that may be contained within a tumor.

While the AmpliconArchitect method used here demonstrates robust ecDNA detection (30), bioinformatic analysis of ecDNA from short-read WGS data still faces several challenges. These include technical hurdles such as detecting structural variations within repetitive genomic regions, uneven sequencing coverage, and the algorithmic complexity of distinguishing different amplicon types, as well as tumorspecific factors like tumor purity and ecDNA CN. Nevertheless, continuous advancements in sequencing technologies and the development of more sophisticated ecDNA detection algorithms are paving the way for more precise estimations of ecDNA frequency and structure in cancer (19).

To conclude, our comprehensive analysis of ecDNA in urothelial carcinoma reveals its early involvement in tumorigenesis, as well as its crucial role in driving the evolution and heterogeneity of multifocal cancer. We also find that ecDNA+ malignant cells express little to no MHC-I molecules, thus enabling evasion of T-cell responses. These results provide novel insights for future ecDNA-targeted therapy research, with potential implications for other ecDNA-driven tumor types as well.

METHODS

CUGA-WGS Datasets, Patients, Samples, and Ethics

We obtained WGS data from 616 tumors, 264 ATs, 195 blood samples, and 167 urine samples collected from 558 patients with urothelial carcinoma as part of the CUGA project. The CUGA project includes 102 patients with urothelial carcinoma from the BIG dataset and 456 patients urothelial carcinoma from the YHD-UC dataset (Supplementary Table S2; refs. 31-33).

In this study, we also included 37 patients diagnosed with urothelial carcinoma who underwent tumor resection at Yantai Yuhuangding (YHD) Hospital (Supplementary Tables S2 and S3). None of the patients received any anticancer treatment before the collection of tumor specimens during surgery. Most patients (35 of 37) used a multiregion tumor sampling approach. For multifocal urothelial tumors, we sampled tumor tissues from various sites. In cases of solitary urothelial tumors, representative tumor regions were collected, spatially separated from each other by a minimum of 5 mm, and distant from the tumor margin. Paired ATs were obtained at least 2 cm away from the tumor margin. Spatially distinct localized tumors, as documented through photography, were promptly collected within 30 minutes after surgery. These samples underwent thorough washing with PBS to remove the blood and were then rapidly snap-frozen in liquid nitrogen. Each surgically resected tumor sample underwent a comprehensive macroscopic examination by 2 to 3 expert genitourinary pathologists. Tumors with low tumor purity were excluded. Patient identifiers were reassigned to ensure anonymity and protect confidentiality.

By integrating data from both the BIG and the YHD datasets, we established a 595-patient cohort (comprising 1,411 whole genomes) for a comprehensive analysis of ecDNA. Detailed clinical and sequencing information is available in Supplementary Tables S3 and S4. This study received approval from the Institutional Review Board (IRB 2021-456; 2022-399; 2022-401; 2022-402) of Yantai Yuhuangding Hospital, and all patients provided written informed consent. This study was conducted under the ethical guidelines of the Helsinki Declaration.

WGS

High-molecular-weight genomic DNA (gDNA) was extracted from tissue samples using MagAttract HMW DNA Kit (Qiagen) following the manufacturer's guidelines. The isolated gDNA underwent a 3-day equilibration period at room temperature for homogenization. DNA concentration was quantified using the Qubit dsDNA BR assay (Invitrogen), whereas DNA integrity was evaluated by electrophoresis on a 0.8% agarose gel. A total of 500 ng of gDNA was utilized as input material. The DNA samples were mechanically sheared to sizes ranging from 300 to 500 bp using the Covaris LE220 instrument. Sequencing libraries were constructed using MGIEasy DNA Library Preparation Kit (MGI) following the manufacturer's protocols. The prepared libraries underwent deep sequencing on the DNBSEQ platform using the PE-150 sequencing strategy.

Uniform Data Processing and Detection of Somatic Variants

To mitigate potential variations in somatic variant detection, all raw sequencing datasets were processed using the same pipelines. The initial quality assessment of FASTQ files was performed using FastQC (v0.11.7; RRID: SCR_014583) from Babraham Bioinformatics. Subsequently, sequencing reads underwent adapter trimming, removal of N content, and elimination of low-quality bases using Fastp (v0.23.2; RRID: SCR_016962; ref. 58). The resulting cleaned FASTQ files were aligned to the human reference genome GRCh38 using BWA-MEM (v0.7.17; RRID: SCR_022192; ref. 59), generating BAM files. BAM files were then sorted and indexed with SAMtools (v1.11; RRID: SCR_002105; ref. 60). Coverage statistics were calculated using SAMtools, with a mean coverage of 41.2× observed for WGS (Supplementary Table S4).

Mutation Calling. Somatic SNVs were identified using Mutect2 (v4.2.5.0; RRID: SCR_000559; ref. 61) with default parameters. Somatic insertions and deletions were detected using both Mutect2 (61) and Strelka2 (v2.9.10; RRID: SCR_005109; ref. 62) with default settings. Each algorithm was executed separately for each tumor sample, utilizing the corresponding paired germline sample as the control. To enhance specificity, insertions and deletions called by two algorithms were retained and then merged using bedtools (v2.30.0; RRID: SCR_006646; ref. 63). The variant call format files were subsequently split and underwent left trimming using GATK (v4.2.5.0; RRID:SCR_001876). ANNOVAR (RRID:SCR_012821; ref. 64) was employed for variant call format file annotation. To mitigate the risk of misidentifying germline variants, we filtered out variants with allele frequencies exceeding 0.01 in the 1000 Genomes database (RRID: SCR_008801; ref. 65) or the Exome Aggregation Consortium (RRID: SCR_004068; ref. 66), unless flagged as pathogenic in the ClinVar (RRID: SCR_006169; ref. 67) database. Tumor mutation burden was

determined by calculating the total count of nonsynonymous mutations divided by the length of exonic regions. Genomic alterations were annotated using the OncoKB (RRID: SCR_014782; ref. 42) annotator tool (https://github.com/oncokb/oncokb-annotator). OncoKB provides curated information about the functional and clinical significance of specific genetic alterations.

CNV Calling. The FACETS (v0.5.14; RRID: SCR_026264; ref. 68), HMMcopy (v1.42.0; RRID: SCR_026464; ref. 69), and ichorCNA (v0.2.0; RRID: SCR_024768; ref. 70) software were used to detect somatic CNVs. FACETS software was utilized to determine the average tumor ploidy and purity, providing allele-specific estimations of CNs. Tumor regions with a major CN \geq 2 fraction exceeding 50% were classified as exhibiting whole-genome duplication events. The fraction of genome altered was defined as the ratio of bases with log₂ CNV (gain or loss) >0.2 to the total number of profiled genome bases. The Genomic Identification of Significant Targets in Cancer (GISTIC, v2.0; RRID: SCR_000151; ref. 71) algorithm was used to assess the CNV status of genes. Genes were categorized as deletion, loss, diploid, gain, or amplification based on GISTIC scores of -2, -1, 0, 1, and 2, respectively. Using GISTIC scores, we computed the intratumoral heterogeneity index for each patient, defined as the proportion of genes with different CNV status across samples relative to all genes. The patient's intratumoral heterogeneity index was averaged across all sample pairs. Eight patients were excluded from intratumoral heterogeneity index calculations due to tumor WGS data originating from different sequencing batches (CUGA-006, 009, 044) or lacking normal control WGS data (CUGA-MR-016, 019, 039, 043, 057). HMMcopy estimated read counts for 500 kbp bins across the genome for five ecDNA+ tumors. The R package ichorCNA was executed with default parameter values to validate obtained CNV profiles.

SV Calling. We utilized four algorithms—Manta (v1.6.0; RRID: SCR_022997; ref. 72), Lumpy (v0.2.13; RRID: SCR_003253; ref. 73), SvABA (v1.1.0; RRID: SCR_022998; ref. 74), and Delly (v1.1.3; RRID: SCR_004603; ref. 75)—to detect SVs in paired tumor–normal WGS data. Each algorithm operated independently. The SV calls generated by each caller were merged using SURVIVOR (v1.0.7; RRID: SCR_022995; ref. 76), allowing 1,000 bp of slop at the breakpoints. SVs called by two algorithms were retained. We computed the number of SVs for each sample. Moreover, the detected SVs were used for further analysis of chromothripsis.

Detection of Clustered Mutations

Kataegis, a process characterized by focal hypermutation and clustered point mutations, was identified using the R package Maftools (v2.14.0; RRID: SCR_024519; ref. 77) by analyzing mutation clusters across the genome. fSCNA-overlapping kataegis was defined as any clustered event that coincided with fSCNA regions.

Drivers

To identify driver genes in urothelial carcinoma, we used three independent methods: 20/20+ (v1.2.3; ref. 78), MutSigCV (v1.41; RRID: SCR_010779; ref. 79), and dNdScv (v0.0.1.0; RRID: SCR_017093; ref. 80). Each software operated independently with default parameters. A total of 37 driver genes were identified by at least two software tools within the CUGA cohort, meeting the criterion of *q* value <0.1 (Supplementary Table S16).

Phylogenetic Tree Construction

To construct the phylogenetic tree of each patient, we used the R package MesKit (v1.1.2; RRID: SCR_020959; ref. 81), using somatic nonsynonymous mutations as input data and utilizing maximum likelihood algorithms.

Analysis of Mutation Signature

The SNVs, including nonsynonymous and synonymous mutations, were classified into 96 substitution types, based on six base substitutions (C > A, C > G, C > T, T > A, T > C, and T > G) and neighboring bases. The MutationalPatterns (v 3.8.1; RRID: SCR_024247; ref. 82) R package was utilized to assess the activity of APOBEC-related mutational signatures (COSMIC SBS2 and SBS13) for each sample.

ecDNA Detection and Characterization

The AmpliconSuite-pipeline wrapper (v1.0.0; RRID: SCR_023150; ref. 30) was performed to identify ecDNA from WGS data. The seed detection pipeline integrated CNVKit (v0.9.9; RRID: SCR_021917; ref. 83) operating in tumor/urine-normal mode to detect somatic CNVs against matched normal WGS samples for each patient. ATs, blood samples, and tumor/urine samples without matched germline samples also underwent the same pipeline in unpaired mode for standalone CNV detection. The CNV calls were processed using the amplified_intervals.py script, filtering regions larger than 50 kbp, and with a CN exceeding 4.5, generating a set of seed regions. The wrapper invoked AmpliconArchitect (v1.3.r6; RRID: SCR_023150; ref. 34) in default mode on the WGS BAM files to examine seed regions and profile the architecture of focal amplifications. The resulting graph and cycles output files were then input into Amplicon-Classifier (v.1.0.0) to classify AmpliconArchitect amplicons, including ecDNA, breakage-fusion-bridge, complex noncyclic, and linear focal amplifications. In cases in which a patient exhibits multiple amplification topologies, we classify them based on the following priority order: ecDNA, breakage-fusion-bridge, complex noncyclic, and linear. We defined ecDNA species based on unique ecDNA sequences within individual samples and determined the number of ecDNA species in each sample by summing the predicted ecDNA species across all detected AA amplicons (30, 84). The number of oscillating CN states of ecDNA was calculated based on the number of CN segments predicated by AmpliconArchitect. AmpliconClassifier also generated BED files corresponding to the classified regions and annotated the identity of genes on the focal amplifications. Candidate ecDNA was visualized using CycleViz (v0.1.5). The similarity score between the two amplicons was calculated as previously described (30), and P < 0.05 was considered significant. Clonal ecDNA was defined as the same ecDNA event detected in all tumor regions from the same patient. Additionally, to reduce false-positive focal amplifications, we manually reviewed all focal amplification outputs and filtered out similar amplicons across different patients, utilizing multiple hypothesis testing corrections with a cutoff of *P* value divided by the total number of patients with focal amplification. The amplicon complexity score was computed based on the Shannon entropy of the distribution of CNs assigned to the amplicon structure decompositions output by AmpliconArchitect (30).

Chromothripsis Detection

Chromothripsis events were evaluated in a cohort of 488 patients with urothelial carcinoma with paired tumor–germline WGS data using ShatterSeek (v1.1; RRID: SCR_026463; ref. 45). Briefly, it uses intrachromosomal SVs initially to identify clusters of interleaved rearrangements. Subsequently, it evaluates various statistical criteria within these regions. The results include a data frame presenting the statistical criteria values and supplementary details per chromosome. Regions potentially indicative of chromothripsis undergo visual scrutiny utilizing local SVs and CN profiles. The chromothripsis events were classified into low-confidence and high-confidence categories following the criteria outlined by Cortés-Ciriano and colleagues (45). Low-confidence calls encompassed 4 to 6 segments between two CN states, whereas high-confidence calls consisted of a minimum of 7 adjacent segments. Both low-confidence and high-confidence calls were retained for further analysis.

Neoantigen Prediction

Class I HLA alleles (HLA-A, HLA-B, and HLA-C) were inferred from WGS data using HLA-Scan (v2.1.4; ref. 85) with default settings. To predict neoantigens associated with identified HLA genotypes, we used the pVACseq (v3.1.2; RRID: SCR_025435; ref. 86) pipeline. The pVACseq pipeline incorporates the NetMHC and NetMHCpan algorithms to identify potential MHC ligands and predict their binding affinities to class I HLA molecules. We specifically considered neoantigens with peptide lengths of 9 or 10 amino acids and binding affinities (IC50) below 500 nmol/L.

Detection of HLA-LOH Events

SpecHLA (v 1.0.7; RRID: SCR_026462; ref. 52) was used to predict LOH events for MHC-I HLA genes from WGS data. LOH events were identified using SpecHLA's default settings.

WES Processing

WES data (Supplementary Table S5) were utilized to detect somatic mutations, using a pipeline identical to that used for WGS. In cases in which both WGS and WES data were available for samples, we integrated the somatic mutation results to examine alterations in driver genes.

Oncogene List

The oncogene list was curated from ONGene (87) and COSMIC (88) databases. The complete gene list (n = 990) is provided in Supplementary Table S13.

RNA-Seq Processing

FastQC was used to evaluate the quality of raw sequencing data. TrimGalore (v0.6.7; RRID: SCR_011847) was then utilized to remove low-quality bases and adapter sequences. The resulting trimmed RNA-seq reads were aligned to the human reference genome GRCh38 using STAR (v2.7.10a; RRID: SCR_004463; ref. 89). RSEM (v1.3.3; RRID: SCR_000262; ref. 90) was used to quantify gene and isoform expression levels from the aligned reads.

Circle-Seq Processing

Circle-seq (Supplementary Table S7) is a sequencing library enrichment protocol utilized for the detection of circular DNAs. Sequencing FASTQ files were converted to BAM files using the same pipeline used for WGS data processing. The ecDNA identified in WGS data underwent visual inspection using the Integrated Genomics Viewer (RRID: SCR_011793; ref. 91).

Timing Analysis of Mutations on Drivers and ecDNA

To infer the relative timing of mutations on driver genes and ecDNA, we categorized SNVs into three groups based on their VAFs: low (VAF \leq 0.333), intermediate (0.333 < VAF \leq 0.667), and high (VAF > 0.667; ref. 9). These groups were interpreted as representing minor, late, and early SNVs within the trajectory of tumor progression (19).

DNA FISH

DNA FISH was performed using HER-2 DNA Probe Kit (FP-001; Healthcare Biotechnology) and CCND1(BCL1)/CEP11 Gene Amplification Probe Detection Kit (FP-041; Healthcare Biotechnology) to confirm ERBB2 and CCND1 amplification status in tissue samples. Initially, FFPE tumor samples underwent deparaffinization in xylene, rehydration through ethanol washes, and rinsing in distilled water. This was followed by heat-induced epitope retrieval and protein digestion. Slides were subsequently dehydrated by washing in 70%, 85%, and 100% cold ethanol stored at -20°C (120 seconds in each solution). FISH probes, diluted in hybridization buffer, were applied to the slides and covered with a coverslip. Denaturation of the slides occurred at 85°C for 5 minutes, followed by overnight (14-18 hours) hybridization at 42°C. Slides were washed with 2×SSC/0.3% NP-40 (pH 7.0-7.5). The dried slides were stained with 10 µL of 4',6-diamidino-2-phenylindole (DAPI) buffer, and images were captured using OLYMPUS BX43.

scRNA-Seq

Data Processing. In-house raw scRNA-seq data (Supplementary Table S8) generated from the DNBelab C4 platform underwent preprocessing steps, including demultiplexing cellular barcodes, read alignment to the hg38 reference genome, and generation of a gene count matrix, using PISA (92). Detailed quality control metrics were generated and assessed, and cells underwent rigorous filtering to ensure high data quality for subsequent analyses. Quality control procedures were performed using the R Seurat (v5.0.3; RRID: SCR_016341; ref. 93) package, involving the exclusion of likely apoptotic or dying cells with a mitochondrial fraction $\geq 25\%$, as well as cells with low-complexity libraries (≤200 genes) or highly complex libraries (≥6,000 genes). Potential doublets were identified and removed using the DoubletFinder (v2.0.3; RRID: SCR_018771; ref. 94) package.

Unsupervised Clustering and Major Cell Type Annotation. The default parameters of Seurat were utilized throughout unless specified otherwise. Briefly, the unique molecular identifier count matrix was normalized using the "NormalizeData" function. Subsequently, 2,000 highly variable genes were identified from the natural log-transformed expression matrix using the "FindVariableFeatures" function with the "vst" method. These variable genes were used for both cell type clustering and dimensionality reduction. After adjusting for unique molecular identifier counts, 30 principal components (PC) were obtained via PC analysis. Initial clustering was performed with the "FindClusters" function using these 30 PCs with a resolution of 0.1. For visualization, Uniform Manifold Approximation and Projection was applied using Seurat's "RunUMAP" function, using the same PCs used for clustering. Six major cell types were identified using known markers: T/NK cells (CD2, CD3D, CD3E, NKG7, and GZMA), B cells (CD79A, CD79B, MS4A1, and MZB1), myeloid cells (CD14, CD163, CD68, and LYZ), epithelial cells (EPCAM, KRT18, KRT19, and KRT7), fibroblasts (COL1A2, COL1A1, ACTA2, and DCN), and endothelial cells (VWF, PECAM1, CDH5, and ENG). Cell clusters expressing markers from multiple major cell types were excluded. Following quality control, 259,808 cells remained for further analysis.

Single-Cell CNV Analysis. We established a reference for normal CNs by randomly selecting 2,000 endothelial cells and 2,000 fibroblasts. Next, we used InferCNV (v1.18.1; RRID: SCR_021140; ref. 49) to identify large-scale CNAs specifically in epithelial cells. Integrating information from cluster distribution and single-cell CNV patterns, we found that all epithelial cells displayed malignant characteristics. Next, for each sample, gene scores were restandardized between -1 and 1. The CNV score for each malignant cell was then calculated as the squared sum of these gene scores. The ComplexHeatmap (v2.14.0; RRID: SCR_017270; ref. 95) R package was used to visualize the pergene CN scores for each malignant cell.

Reclustering of Malignant Cells and Immune Cells. To analyze malignant cells and three immune cell subgroups, we corrected for patient-specific variations using batch correction with the R package harmony (v1.2.0; RRID: SCR_022206; ref. 96). Dimensionality reduction and unsupervised clustering parameters were set to align with major cell types, except for resolution, which was adjusted to 1 for finer clustering (0.5 for malignant cells). Seurat's "FindAllMarkers" function was used to pinpoint differentially expressed genes (DEG) within each cluster, using default parameters (min.pct = 0.25, logfc. threshold = 0.25). Subsequently, immune subtypes were identified by leveraging known published signatures, and these identified DEGs. T cells included CD4⁺ T cells, CD8⁺ T cells, Tregs (*FOXP3, IL2RA*, and *IKZF2*), naïve T cells (*LEF1* and *TCF7*), and IFN⁺ T cells (*IFIT1* and *ISG15*). NK cells were further categorized into CD56^{dim}CD16^{hi} and CD56^{bright}CD16^{lo} subtypes based on the expression levels of CD16 (*FCGR3A*) and CD56 (*NCAM1*). Among B cells, three subtypes were identified: naïve B cells (*FCER2* and *TCL1A*), memory B cells (*AIM2* and *TNFRSF13B*), and plasma cells (*IGKC* and *IGHG3*). Myeloid cells were divided into macrophages (*CD68, CD163,* and *C1QA*), conventional dendritic cells of subtypes 1 (*CLEC9A* and *XCR1*), 2 (*FCER1A* and *CD1C*), and 3 (*LAMP3* and *CCL19*), mast cells (*TPSAB1* and *TPSB2*), and monocytes (*FCN1, VCAN*, and *S100A8*).

Gene Signature Scores. Multiple gene signature scores were calculated based on the scRNA-seq data using the "AddModuleScore" function of Seurat. For malignant cells, the APM score was calculated based on the expression of six genes (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, *HLA-F*, and *B2M*).

Expression Programs of Intratumoral Heterogeneity. To systematically characterize transcriptional heterogeneity, we performed unsupervised NMF using the cnmf Python package (v1.4.1; RRID: SCR_025495; ref. 97) on the 13 tumors containing more than 100 malignant cells. The NMF method identified sets of coexpressed genes, termed metagenes. To determine the optimal number of metagenes (K) for each tumor, we balanced stability and reconstruction error within the data. A total of 50 programs were identified across the 13 tumors. The top-ranked genes for each program, based on their NMF factor loadings, are detailed in Supplementary Table S19. Subsequently, we performed hierarchical clustering on these 50 programs, utilizing one minus the Pearson correlation coefficient as the distance metric across all gene scores. This analysis led to the identification of five metaprograms. We leveraged the HALLMARK (98) and CancerSEA (99) collections to functionally annotate both programs and metaprograms. For each metaprogram signature, we combined the top 50 genes from each constituent metagene. We then calculated the average loading for each gene. Repetitive genes retained their original loadings, whereas the loadings for exclusive genes were summed. Finally, these combined loadings were normalized by dividing by the number of metagenes within the signature. We designated the top 30 genes with the highest resulting values as marker genes for the signature. The metaprogram score of each malignant cell was subsequently computed using the "AddModuleScore" function from the Seurat package.

Pathway Analysis. To identify DEGs between malignant cells of ecDNA⁻ and ecDNA⁺ tumors, we used a two-pronged approach. First, we utilized the "FindMarkers" function within the Seurat package to identify significantly altered genes at the single-cell level (P value < 0.05). Second, we leveraged the pseudobulk method implemented in the Libra (v1.0.0; RRID: SCR_016608; ref. 100) package to analyze gene expression changes at the bulk level (P value < 0.01). Genes identified as significant by both methods were considered bona fide DEGs. Subsequently, we performed pathway enrichment analysis to elucidate the functional roles of these DEGs. Pathway and gene set collections were obtained from publicly available resources: the Molecular Signatures Database HALLMARK (98) and CancerSEA (99). We used the "enricher" R function (default parameters) of the clusterProfiler (v4.6.2; RRID: SCR_016884; ref. 101) R package to assess the over- and under-expression of gene sets within the identified DEGs, leveraging the hypergeometric distribution for enrichment P value calculations. Gene identifiers were mapped using the org.Hs.eg.db (v3.16.0) package from AnnotationDbi. We used the "enricher" R function (default parameters). The top seven enriched pathways (ranked by *P* value for each group) were then visualized as bar plots using the ggplot2 (v3.5.0; RRID: SCR_014601) R packages.

Cell-to-Cell Communication. We used CellChat (v1.6.1; RRID: SCR_021946; ref. 102) to elucidate potential intercellular communication networks. We first imported the normalized expression matrix generated by Seurat and constructed a CellChat object using the "createCellChat" function. Next, we used the "identifyOverExpressedGenes" and "identifyOverExpressedInteractions" functions within CellChat to identify significantly overexpressed ligands and receptors, respectively. These molecules potentially mediate intercellular signaling. We then quantified the likelihood of ligand-receptor interactions using the "computeCommunProb" and "computeCommunProbPathway" functions. Finally, the "aggregateNet" function was utilized to construct the aggregated cell-cell communication network, summarizing these potential communication events.

IHC and mIF Staining

FFPE tumor tissue blocks were cut into thin sections (4-8 µm) and deparaffinized with xylene. To prepare for antibody staining, the sections were rehydrated through a graded ethanol series (100%, 95%, and 70%) followed by microwave treatment in the citric acid solution for antigen retrieval (15 minutes). IHC was performed using a CD45 primary antibody (Cat# ZM-0183, ZSGB-BIO, RRID: AB_3676256) and evaluated by two independent pathologists. We performed mIF staining using two antibody panels: panel 1 [CD20 (Cell Signaling Technology, #48750S, RRID:AB_3107071), CD68 (Abcam, #ab213363, RRID:AB_2801637), CD4 (Zenbio, #R50028, RRID:AB_3676254), and FOXP3 (Abcam, #ab20034, RRID:AB_445284)] and panel 2 [CD8a (Cell Signaling Technology, #85336, RRID:AB_2800052), pan-CK (Abcam, #308262, RRID:AB_3676255), and HLA-I (Proteintech, #15240, RRID:AB_1557426)]. Akoya OPAL Polaris 7-Color Automation IHC Kit (NEL871001KT, RRID: AB_3674065) was used following the manufacturer's instructions. Briefly, FFPE slides were incubated sequentially with primary antibodies. Secondary antibodies conjugated to Opal fluorophores were then applied. Finally, nuclei were stained with DAPI, and multispectral images were acquired using a PANNORAMIC SCAN II scanner. Images were processed using QuPath (v0.5.1; ref. 103).

Deconvolution of Bulk RNA-Seq for Immune Cell Estimation

We deconvoluted bulk RNA-seq data from 331 tumor samples to estimate the relative proportions of immune cell types. Utilizing the build_model function within the omnideconv (v0.0.0.9000) package (https://github.com/omnideconv/omnideconv), we constructed a reference signature matrix. This matrix encompasses the top 20 marker genes for each major cell type identified in our scRNAseq data, as well as pan-marker genes (Supplementary Table S23). CIBERSORTx (RRID: SCR_016955; ref. 104) then leveraged this matrix to estimate the relative abundance of each cell type within each sample.

ISTAT Analysis

The BED files for focal amplification regions were obtained from AmpliconClassifier. To evaluate the statistical significance of the overlap between two genome annotations (ecDNA and other fSCNAs; urine ecDNA and tumor ecDNA), we used the ISTAT software (v1.0; https://github.com/shahab-sarmashghi/ISTAT; ref. 35).

Statistical Analysis

All statistical analyses were performed in R (v4.3.0). The statistical tests used to analyze the data were described in figure legends and the results section. In multivariate Cox proportional hazards modeling,

we first assessed the impact of six clinical features (age, sex, smoking history, tumor stage, grade, and tumor location) and three fSCNA subtypes on overall survival using univariate analyses. Three clinical features (age, grade, and tumor stage) and fSCNA subtypes demonstrated significant associations with survival. To address potential multicollinearity and prevent overfitting in subsequent multivariate analyses, we used a stepwise backward selection procedure using the Akaike information criterion (AIC). For categorical variables, pTa/ pT1 and tumors without fSCNA were used as reference groups.

In our logistic regression analysis, continuous variables were dichotomized into high and low groups based on their median values. Prevalence estimates were accompanied by 95% confidence intervals (CI) calculated using the propCI function. Logistic regression models were fitted using the glm function (family = "binomial"), and OR with 95% CIs were reported.

Data Availability

A detailed overview of data accessibility, including data origins and accession identifiers for all sequencing datasets, is provided in Supplementary Tables S4-S8. The raw WGS data, histology, and metadata of the BIG cohort can be acquired from the Genome Sequence Archive for Human with accession numbers HRA001867 and HRA000029. All sequencing data from the YHD dataset is available in the Genome Sequence Archive for Human with accession numbers HRA003461, HRA005001, HRA004718, and HRA005963. The raw genome sequencing data are protected; deidentified data are available under restricted access to protect patient privacy and comply with the Regulations on Management of Human Genetic Resources in China. These data can be requested for research use from the corresponding author. The identified ecDNA sequences in the CUGA dataset can be visualized and further analyzed on a dedicated web platform: AmpliconRepository (ampliconrepository.org/project/CUGA). The ecDNA profiles of multiple cancer types in the The Cancer Genome Atlas and Pan-Cancer Analysis of Whole Genomes datasets can also be acquired from AmpliconRepository. All other remaining data are available within the article and Supplementary Data. Any additional information required to reanalyze the data reported in this article is available upon request. No standalone software or code was generated to analyze the data. All software and analysis codes are publicly available. The tools and versions used are detailed in the Methods and Supplementary Table S24. The detailed codes and instructions for all software have been deposited on GitHub (https://github.com/ DreamLab-WeiLv/CUGA-ecDNA).

Ethics Reporting

All patients provided written informed consent per the Declaration of Helsinki principles.

Authors' Disclosures

R.G.W. Verhaak reports other support from Boundless Bio during the conduct of the study. B.M. Faltas reports Advisory/Scientific Board Member: Bladder Cancer Advocacy Network, Inc., Other Interest: Guardant Health, Inc., and Professional Services: UroToday. S. Wu is a member of the SAB of Dimension Genomics. P.S. Mischel reports personal fees and other support from Boundless Bio outside the submitted work. A.G. Henssen reports personal fees from Econic Biosciences outside the submitted work. V. Bafna reports grants from Cancer Research UK, the NCI, and the NIGMS during the conduct of the study, as well as other support from Boundless Bio, Inc. outside the submitted work; in addition, V. Bafna is a cofounder, serves on the scientific advisory board of Boundless Bio, Inc., and Abterra Inc., and holds equity in both companies. B. Regenberg reports B. Regenberg is cofounder of the company CARE-DNA IP ApS (CVR-number 45303667). No disclosures were reported by the other authors.

Authors' Contributions

W. Lv: Conceptualization, data curation, formal analysis, supervision, validation, investigation, visualization, methodology, writingoriginal draft, project administration, writing-review and editing. Y. Zeng: Data curation, formal analysis, validation, investigation, visualization. C. Li: Data curation, formal analysis, validation, investigation, visualization, methodology, writing-review and editing. Y. Liang: Resources, data curation, formal analysis, investigation, visualization, writing-review and editing. H. Tao: Data curation, formal analysis, investigation. Y. Zhu: Investigation, writing-review and editing. X. Sui: Investigation. Yue Li: Validation, investigation. S. Jiang: Validation, investigation. Q. Gao: Validation, investigation. E. Rodriguez-Fos: Investigation. G. Prasad: Software, investigation. Y. Wang: Investigation. R. Zhou: Investigation. Z. Xu: Investigation. X. Pan: Investigation. L. Chen: Investigation. X. Xiang: Investigation. H. Teng: Investigation. C. Sun: Investigation. T. Qin: Investigation. W. Dong: Resources, investigation. Yongwei Li: Investigation. X. Lan: Investigation. X. Li: Investigation. L. Lin: Investigation. L. Bolund: Resources, investigation. H. Yang: Resources, investigation. R.G.W. Verhaak: Investigation, writing-review and editing. B.M. Faltas: Investigation. J.B. Hansen: Investigation, writing-review and editing. S. Wu: Investigation, visualization, writing-review and editing. P.S. Mischel: Investigation, writing-review and editing. A.G. Henssen: Investigation. V. Bafna: Software, investigation, methodology, writing-review and editing. J. Luebeck: Software, investigation, methodology, writing-review and editing. B. Regenberg: Conceptualization, formal analysis, supervision, investigation, writing-review and editing. Y. Luo: Conceptualization, resources, supervision, investigation, writing-review and editing. C. Lin: Conceptualization, resources, formal analysis, supervision, funding acquisition, investigation, writing-review and editing. P. Han: Conceptualization, resources, supervision, investigation, project administration, writing-review and editing.

Acknowledgments

We thank the patients and their families for their participation in the individual projects. We gratefully acknowledge the contributions of the The Cancer Genome Atlas and Pan-Cancer Analysis of Whole Genomes consortiums. This work was funded by grants from the IBMC-BGI Center, the Taishan Scholar Program (No. Tsqn202103198), and the National Nature Science Foundation of China (No. 82270722, and No. 32200421). S.W. is a scholar of and is supported by the Cancer Prevention and Research Institute of Texas (RR210034). B. Regenberg was supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement number 899417, NNF21OC0072023, and Sygeforsikringen Danmark (2021-0321). P.S. Mischel is supported by the Cancer Grand Challenges partnership financed by Cancer Research UK (CG-CATF-2021/100025) and the NCI (OT2CA278688). P.S. Mischel is the eDyNAmiC team lead, and R.G.W. Verhaak, S. Wu, A.G. Henssen, and V. Bafna are members of the eDyNAmiC team.

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

Received October 20, 2024; revised January 20, 2025; accepted March 6, 2025; posted first March 7, 2025.

REFERENCES

1. Paner GP, Smith SC, Hartmann A, Agarwal PK, Compérat E, Amin MB. Flat intraurothelial lesions of the urinary bladder-do hyperplasia, dysplasia, and atypia of unknown significance need to exist as



diagnostic entities? and how to handle in routine clinical practice. Mod Pathol 2022;35:1296-305.

- 2. Czerniak B, Dinney C, McConkey D. Origins of bladder cancer. Annu Rev Pathol 2016;11:149–74.
- 3. Dyrskjøt L, Hansel DE, Efstathiou JA, Knowles MA, Galsky MD, Teoh J, et al. Bladder cancer. Nat Rev Dis Primers 2023;9:58.
- Knowles MA, Hurst CD. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nat Rev Cancer 2015;15:25–41.
- Tran L, Xiao JF, Agarwal N, Duex JE, Theodorescu D. Advances in bladder cancer biology and therapy. Nat Rev Cancer 2021;21: 104–21.
- Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. Cell 2017;171:540–56.e25.
- Steele CD, Abbasi A, Islam SMA, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. Nature 2022;606:984–91.
- Fujii Y, Sato Y, Suzuki H, Kakiuchi N, Yoshizato T, Lenis AT, et al. Molecular classification and diagnostics of upper urinary tract urothelial carcinoma. Cancer Cell 2021;39:793–809.e8.
- 9. Nguyen DD, Hooper WF, Liu W, Chu TR, Geiger H, Shelton JM, et al. The interplay of mutagenesis and ecDNA shapes urothelial cancer evolution. Nature 2024;635:219–28.
- Lv W, Pan X, Han P, Wu S, Zeng Y, Wang Q, et al. Extrachromosomal circular DNA orchestrates genome heterogeneity in urothelial bladder carcinoma. Theranostics 2024;14:5102–22.
- 11. Kim H, Nguyen NP, Turner K, Wu S, Gujar AD, Luebeck J, et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. Nat Genet 2020;52: 891–7.
- Wu S, Turner KM, Nguyen N, Raviram R, Erb M, Santini J, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. Nature 2019;575:699–703.
- Cox D, Yuncken C, Spriggs AI. Minute chromatin bodies in malignant tumours of childhood. Lancet 1965;1:55–8.
- Noer JB, Hørsdal OK, Xiang X, Luo Y, Regenberg B. Extrachromosomal circular DNA in cancer: history, current knowledge, and methods. Trends Genet 2022;38:766–81.
- Yi E, Chamorro González R, Henssen AG, Verhaak RGW. Extrachromosomal DNA amplifications in cancer. Nat Rev Genet 2022;23: 760–71.
- Pang J, Nguyen N, Luebeck J, Ball L, Finegersh A, Ren S, et al. Extrachromosomal DNA in HPV-mediated oropharyngeal cancer drives diverse oncogene transcription. Clin Cancer Res 2021;27:6772–86.
- Pongor LS, Schultz CW, Rinaldi L, Wangsa D, Redon CE, Takahashi N, et al. Extrachromosomal DNA amplification contributes to small cell lung cancer heterogeneity and is associated with worse outcomes. Cancer Discov 2023;13:928–49.
- Chapman OS, Luebeck J, Sridhar S, Wong IT, Dixit D, Wang S, et al. Circular extrachromosomal DNA promotes tumor heterogeneity in high-risk medulloblastoma. Nat Genet 2023;55:2189–99.
- Bailey C, Pich O, Thol K, Watkins TBK, Luebeck J, Rowan A, et al. Origins and impact of extrachromosomal DNA. Nature 2024;635: 193–200.
- Kim H, Kim S, Wade T, Yeo E, Lipsa A, Golebiewska A, et al. Mapping extrachromosomal DNA amplifications during cancer progression. Nat Genet 2024;56:2447–54.
- 21. Yi E, Gujar AD, Guthrie M, Kim H, Zhao D, Johnson KC, et al. Livecell imaging shows uneven segregation of extrachromosomal DNA elements and transcriptionally active extrachromosomal DNA hubs in cancer. Cancer Discov 2022;12:468–83.
- 22. deCarvalho AC, Kim H, Poisson LM, Winn ME, Mueller C, Cherba D, et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. Nat Genet 2018;50:708–17.
- Nathanson DA, Gini B, Mottahedeh J, Visnyei K, Koga T, Gomez G, et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. Science 2014;343:72–6.

- Chen L, Zhang C, Xue R, Liu M, Bai J, Bao J, et al. Deep whole-genome analysis of 494 hepatocellular carcinomas. Nature 2024;627: 586–93.
- Koche RP, Rodriguez-Fos E, Helmsauer K, Burkert M, MacArthur IC, Maag J, et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. Nat Genet 2020;52:29–34.
- Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature 2017;543:122–5.
- Xu K, Ding L, Chang TC, Shao Y, Chiang J, Mulder H, et al. Structure and evolution of double minutes in diagnosis and relapse brain tumors. Acta Neuropathol 2019;137:123–37.
- Chamorro González R, Conrad T, Stöber MC, Xu R, Giurgiu M, Rodriguez-Fos E, et al. Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells. Nat Genet 2023;55:880–90.
- Li R, Du Y, Chen Z, Xu D, Lin T, Jin S, et al. Macroscopic somatic clonal expansion in morphologically normal human urothelium. Science 2020;370:82–9.
- Luebeck J, Ng AWT, Galipeau PC, Li X, Sanchez CA, Katz-Summercorn AC, et al. Extrachromosomal DNA in the cancerous transformation of Barrett's oesophagus. Nature 2023;616:798–805.
- Lu H, Liang Y, Guan B, Shi Y, Gong Y, Li J, et al. Aristolochic acid mutational signature defines the low-risk subtype in upper tract urothelial carcinoma. Theranostics 2020;10:4323–33.
- 32. Liang Y, Tan Y, Guan B, Guo B, Xia M, Li J, et al. Single-cell atlases link macrophages and CD8⁺ T-cell subpopulations to disease progression and immunotherapy response in urothelial carcinoma. Theranostics 2022;12:7745–59.
- Zeng Y, Lv W, Tao H, Li C, Jiang S, Liang Y, et al. Mapping the chromothripsis landscape in urothelial carcinoma unravels great intratumoral and intertumoral heterogeneity. iScience 2025;28:111510.
- Deshpande V, Luebeck J, Nguyen NPD, Bakhtiari M, Turner KM, Schwab R, et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. Nat Commun 2019;10:392.
- Sarmashghi S, Bafna V. Computing the statistical significance of overlap between genome annotations with iStat. Cell Syst 2019;8: 523–9.e4.
- Bergstrom EN, Luebeck J, Petljak M, Khandekar A, Barnes M, Zhang T, et al. Mapping clustered mutations in cancer reveals APOBEC3 mutagenesis of ecDNA. Nature 2022;602:510–7.
- Hadi K, Yao X, Behr JM, Deshpande A, Xanthopoulakis C, Tian H, et al. Distinct classes of complex structural variation uncovered across thousands of cancer genome graphs. Cell 2020;183:197–210.e32.
- Lange JT, Rose JC, Chen CY, Pichugin Y, Xie L, Tang J, et al. The evolutionary dynamics of extrachromosomal DNA in human cancers. Nat Genet 2022;54:1527–33.
- 39. Purshouse K, Friman ET, Boyle S, Dewari PS, Grant V, Hamdan A, et al. Oncogene expression from extrachromosomal DNA is driven by copy number amplification and does not require spatial clustering in glioblastoma stem cells. Elife 2022;11:e80207.
- Aran D, Camarda R, Odegaard J, Paik H, Oskotsky B, Krings G, et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat Commun 2017;8:1077.
- Greaves M, Maley CC. Clonal evolution in cancer. Nature 2012; 481:306–13.
- Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. JCO Precision Oncol 2017;2017:1-16.
- Eischen CM. Genome stability requires p53. Cold Spring Harb Perspect Med 2016;6:a026096.
- 44. Hanel W, Moll UM. Links between mutant p53 and genomic instability. J Cell Biochem 2012;113:433–9.
- 45. Cortés-Ciriano I, Lee JJ, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet 2020;52:331–41.
- Debatin NF, Bady E, Mandelkow T, Huang Z, Lurati MCJ, Raedler JB, et al. Prognostic impact and spatial interplay of immune cells in urothelial cancer. Eur Urol 2024;86:42–51.

- 47. Lin MS, Jo S-Y, Luebeck J, Chang HY, Wu S, Mischel PS, et al. Transcriptional immune suppression and up-regulation of doublestranded DNA damage and repair repertoires in ecDNA-containing tumors. eLife 2024;12:RP88895.
- 48. Wu T, Wu C, Zhao X, Wang G, Ning W, Tao Z, et al. Extrachromosomal DNA formation enables tumor immune escape potentially through regulating antigen presentation gene expression. Sci Rep 2022;12:3590.
- Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 2014;344:1396–401.
- Tay C, Tanaka A, Sakaguchi S. Tumor-infiltrating regulatory T cells as targets of cancer immunotherapy. Cancer Cell 2023;41: 450-65.
- Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, et al. Immune evasion in cancer: mechanistic basis and therapeutic strategies. Semin Cancer Biol 2015;35(Suppl):S185–s98.
- Wang S, Wang M, Chen L, Pan G, Wang Y, Li SC. SpecHLA enables full-resolution HLA typing from sequencing data. Cell Rep Methods 2023;3:100589.
- 53. Lv W, Pan X, Han P, Wang Z, Feng W, Xing X, et al. Circle-Seq reveals genomic and disease-specific hallmarks in urinary cell-free extrachromosomal circular DNAs. Clin Transl Med 2022;12:e817.
- Zhu Y, Gujar AD, Wong CH, Tjong H, Ngan CY, Gong L, et al. Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. Cancer Cell 2021;39: 694–707.e7.
- Tang J, Weiser NE, Wang G, Chowdhry S, Curtis EJ, Zhao Y, et al. Enhancing transcription-replication conflict targets ecDNA-positive cancers. Nature 2024;635:210–8.
- Pradella D, Zhang M, Gao R, Yao MA, Gluchowska KM, Cendon-Florez Y, et al. Engineered extrachromosomal oncogene amplifications promote tumorigenesis. Nature 2025;637:955–64.
- Zeng Y, Wang A, Lv W, Wang Q, Jiang S, Pan X, et al. Recent development of urinary biomarkers for bladder cancer diagnosis and monitoring. Clin Transl Discov 2023;3:e183.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589–95.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–9.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213–9.
- 62. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods 2018;15:591–4.
- 63. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
- 65. 1000 Genomes Project Consortium; Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–91.
- 67. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. Nucleic Acids Res 2020;48:D835-d44.
- Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res 2016;44:e131.
- 69. Ha G, Roth A, Lai D, Bashashati A, Ding J, Goya R, et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic

expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome Res 2012;22:1995–2007.

- Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat Commun 2017;8:1324.
- 71. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12:R41.
- Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016;32:1220–2.
- 73. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 2014;15:R84.
- 74. Wala JA, Bandopadhayay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, et al. SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res 2018;28: 581–91.
- Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 2012;28:i333–9.
- 76. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun 2017;8:14061.
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747–56.
- Tokheim CJ, Papadopoulos N, Kinzler KW, Vogelstein B, Karchin R. Evaluating the evaluation of cancer driver genes. Proc Natl Acad Sci U S A 2016;113:14330–5.
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013;499:214–8.
- Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. Cell 2017;171:1029–41.e21.
- Liu M, Chen J, Wang X, Wang C, Zhang X, Xie Y, et al. MesKit: a tool kit for dissecting cancer evolution of multi-region tumor biopsies through somatic alterations. Gigascience 2021;10:giab036.
- Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med 2018;10:33.
- Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol 2016;12:e1004873.
- Hung KL, Jones MG, Wong IT-L, Curtis EJ, Lange JT, He BJ, et al. Coordinated inheritance of extrachromosomal DNAs in cancer cells. Nature 2024;635:201–9.
- Ka S, Lee S, Hong J, Cho Y, Sung J, Kim HN, et al. HLAscan: genotyping of the HLA region using next-generation sequencing data. BMC Bioinformatics 2017;18:258.
- Hundal J, Kiwala S, McMichael J, Miller CA, Xia H, Wollam AT, et al. pVACtools: a computational toolkit to identify and visualize cancer neoantigens. Cancer Immunol Res 2020;8:409–20.
- Liu Y, Sun J, Zhao M. ONGene: a literature-based database for human oncogenes. J Genet Genomics 2017;44:119–21.
- Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res 2019;47:D941–7.
- 89. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics 2011;12:323.

- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29: 24–6.
- 92. Shi Q, Liu S, Kristiansen K, Liu L. The FASTQ+ format and PISA. Bioinformatics 2022;38:4639-42.
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36:411–20.
- McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst 2019;8:329–37.e4.
- 95. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847-9.
- Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16:1289–96.
- Kotliar D, Veres A, Nagy MA, Tabrizi S, Hodis E, Melton DA, et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-seq. Elife 2019;8:e43803.

- Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1:417–25.
- 99. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, et al. CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res 2019;47:D900–d8.
- Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nat Commun 2021;12:5692.
- 101. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb) 2021;2:100141.
- 102. Jin S, Guerrero-Juarez CF, Zhang L, Chang J, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using Cell-Chat. Nat Commun 2021;12:1088.
- 103. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7:16878.
- 104. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;37:773–82.