

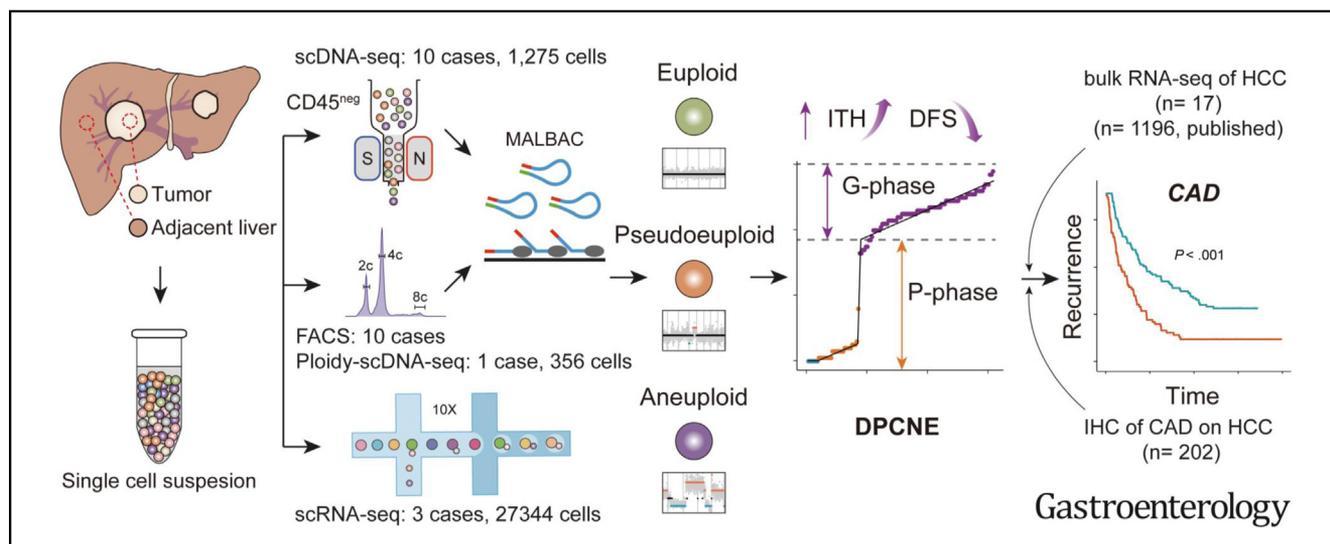
# BASIC AND TRANSLATIONAL—LIVER

## Single-Cell DNA Sequencing Reveals Punctuated and Gradual Clonal Evolution in Hepatocellular Carcinoma



Lin Guo,<sup>1,2</sup> Xianfu Yi,<sup>3</sup> Lu Chen,<sup>1,4</sup> Ti Zhang,<sup>4</sup> Hua Guo,<sup>1</sup> Ziye Chen,<sup>1</sup> Jinghui Cheng,<sup>5</sup> Qi Cao,<sup>5</sup> Hengkang Liu,<sup>5</sup> Chunyu Hou,<sup>5</sup> Lisha Qi,<sup>6</sup> Zhiyan Zhu,<sup>7</sup> Yucun Liu,<sup>8</sup> Ruirui Kong,<sup>5</sup> Chong Zhang,<sup>9</sup> Xiaohua Zhou,<sup>10</sup> Zemin Zhang,<sup>9</sup> Tianqiang Song,<sup>4</sup> Ruidong Xue,<sup>5</sup> and Ning Zhang<sup>1,5</sup>

<sup>1</sup>Laboratory of Tumor Cell Biology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China; <sup>2</sup>Department of Genetics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China; <sup>3</sup>School of Biomedical Engineering and Technology, Department of Bioinformatics, The Province and Ministry Co-Sponsored Collaborative Innovation Center for Medical Epigenetics, Tianjin Medical University, Tianjin, China; <sup>4</sup>Department of Hepatobiliary Cancer, Liver Cancer Research Center, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, China; <sup>5</sup>Translational Cancer Research Center, Peking University First Hospital, Beijing, China; <sup>6</sup>Department of Pathology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China; <sup>7</sup>Tianjin Research Center of Basic Medical Science, Tianjin Medical University, Tianjin, China; <sup>8</sup>Division of General Surgery, Peking University First Hospital, Beijing, China; <sup>9</sup>Beijing International Center for Mathematical Research, Peking University, Beijing, China; and <sup>10</sup>Biomedical Pioneering Innovation Center, Beijing Advanced Innovation Center for Genomics, and School of Life Sciences, Peking University, Beijing, China



See editorial on page 46.

**BACKGROUND & AIMS:** Copy number alterations (CNAs), elicited by genome instability, are a major source of intratumor heterogeneity. How CNAs evolve in hepatocellular carcinoma (HCC) remains unknown. **METHODS:** We performed single-cell DNA sequencing (scDNA-seq) on 1275 cells isolated from 10 patients with HCC, ploidy-resolved scDNA-seq on 356 cells from 1 additional patient, and single-cell RNA sequencing on 27,344 cells from 3 additional patients. Three statistical fitting models were compared to investigate the CNA accumulation pattern. **RESULTS:** Cells in the tumor were categorized into the following 3 subpopulations: euploid, pseudo-euploid, and aneuploid. Our scDNA-seq analysis revealed that CNA accumulation followed a dual-phase copy number evolution model, that is, a punctuated phase followed by a gradual phase. Patients who exhibited

prolonged gradual phase showed higher intratumor heterogeneity and worse disease-free survival. Integrating bulk RNA sequencing of 17 patients with HCC, published datasets of 1196 liver tumors, and immunohistochemical staining of 202 HCC tumors, we found that high expression of *CAD*, a gene involved in pyrimidine synthesis, was correlated with rapid tumorigenesis and reduced survival. The dual-phase copy number evolution model was validated by our single-cell RNA sequencing data and published scDNA-seq datasets of other cancer types. Furthermore, ploidy-resolved scDNA-seq revealed the common clonal origin of diploid- and polyploid-aneuploid cells, suggesting that polyploid tumor cells were generated by whole genome doubling of diploid tumor cells. **CONCLUSIONS:** Our work revealed a novel dual-phase copy number evolution model, showed HCC with longer gradual phase was more severe, identified *CAD* as a promising biomarker for early recurrence of HCC, and supported the diploid origin of polyploid HCC.

**Keywords:** Liver Cancer; Ploidy Heterogeneity; Clonal Evolution; Early Recurrence.

Hepatocellular carcinoma (HCC) accounts for approximately 80% of primary liver cancer and is the third leading cause of cancer-related mortality worldwide.<sup>1</sup> Current targeted therapies and immunotherapies show only limited efficacy and response rate in HCC,<sup>2,3</sup> partially due to extensive tumor heterogeneity.<sup>4,5</sup> We, and others, have characterized intratumor heterogeneity (ITH) in HCC, highlighting the molecular barrier for accurate diagnosis and effective treatment.<sup>6–9</sup> Nevertheless, most studies analyze bulk tumors, which only reflect average profiles of different subclones. The heterogeneity landscape of HCC tumor cells at single-cell resolution remains largely unknown.

Copy number alterations (CNAs), elicited by genome instability, are a major source of ITH.<sup>10</sup> Indeed, a recent pan-cancer analysis indicated that CNAs are a major driver for transcriptomic ITH.<sup>11</sup> Single-cell DNA sequencing (scDNA-seq) can accurately determine CNAs, making it a powerful tool to dissect the complex tumor subclonal structure.<sup>12</sup> Previous scDNA-seq studies of HCC analyzed a total of only 100 tumor cells.<sup>13,14</sup> Although single-cell RNA sequencing (scRNA-seq) studies of HCC revealed the transcriptomic ITH across 40,000 cells, these studies lack genomic resolution.<sup>9,15</sup> A comprehensive scDNA-seq study of HCC tumors is still needed.

Whether tumor progression follows gradual or punctuated evolution is intensely debated.<sup>16–18</sup> Gradual evolution means that tumor cells gradually accumulate genetic aberrations, continually adapt to the selection pressure, and successively transform into more malignant states.<sup>19</sup> By contrast, punctuated evolution model means that genetic alterations are acquired in a short burst of genomic crisis, followed by stable clonal expansion.<sup>20,21</sup> Gao et al<sup>22</sup> performed scDNA-seq on 1000 cells from triple-negative breast cancer (TNBC) and revealed limited ITH across these tumors. They proposed punctuated copy number evolution (PCNE) and suggested that CNAs were acquired in a short burst of genomic crisis, followed by stable clonal expansion. However, our previous scDNA-seq analyses of a liver tumor<sup>23</sup> and of circulating tumor cells<sup>24</sup> showed extensive CNA heterogeneity, which did not fit with the PCNE model. Therefore, we speculated that the CNA evolution in HCC might follow a different model.

Here, we report scDNA-seq on a total of 1631 cells from 11 HCC cases and scRNA-seq on 27,344 cells from 3 additional HCC cases. We propose a novel model of dual-phase copy number evolution (DPCNE) in HCC, which also applies to other cancer types. Based on this model, we identified that *CAD*, a novel candidate gene involved in pyrimidine synthesis, correlates with prolonged gradual phase, rapid tumorigenesis, and reduced survival. Intriguingly, our ploidy-scDNA-seq on 356 cells covering 5 different ploidy subsets from a single case supports the diploid origin of polyploid HCC tumors. Our study presents a unique approach to understanding HCC formation,

## WHAT YOU NEED TO KNOW

### BACKGROUND AND CONTEXT

The heterogeneity landscape and evolution pattern of CNAs in HCC at single-cell resolution remain unknown.

### NEW FINDINGS

CNA accumulation in HCC follows a novel DPCNE model. Prolonged gradual phase correlates with higher ITH and worse prognosis. *CAD* is a biomarker for early recurrence. Polyploid tumor cells have a diploid origin.

### LIMITATIONS

Only 1 case of scDNA-seq in this study has ploidy resolution. Further studies are needed to fully address ploidy heterogeneity in HCC.

### IMPACT

Our DPCNE model provides new perspective on tumor clonal evolution and can be used to identify candidate genes that might be exploited as biomarkers or therapies.

detection, and severity, providing important biological and clinical implications for HCC and shedding new light on tumor evolution in general.

## Materials and Methods

Details about patient cohort, single-cell isolation, flow sorting of diploid and polyploid subsets, single-cell DNA sequencing, integer copy number calculation, copy number profiles clustering, multiple-cell segmentation and event matrix construction, annotation of cancer-related genes in CNAs, classification of cell subpopulations, maximum-parsimony trees, statistical model fitting, classification of P-group and G-group, cell-ITH, intratumor and intertumor heterogeneity, bulk RNA-seq, scRNA-seq, inferring CNAs from scRNA-seq data, survival analysis, and immunohistochemical staining are described in the [Supplemental Material](#).

## Results

### Single-Cell DNA Sequencing Identified 3 Subpopulations of Cells in the Tumor Tissue

We performed scDNA-seq on a total of 1222 cells from tumor tissues (range, 40–182; median, 142) and 53 cells (range, 3–10; median, 4.5) from adjacent nontumorous liver (AL) tissues, respectively ([Figure 1A](#), [Supplementary Tables 1 and 2](#)). To avoid sampling biases introduced by

**Abbreviations used in this paper:** AL, adjacent nontumorous liver; CNA, copy number alteration; DFS, disease-free survival; DPCNE, dual-phase copy number evolution; FACS, fluorescence-activated cell sorting; GCNE, gradual copy number evolution; HCC, hepatocellular carcinoma; IHC, immunohistochemistry; ITH, intratumor heterogeneity; OS, overall survival; PCNE, punctuated copy number evolution; PWD-CNA, pairwise distances of copy number alteration profile; scDNA-seq, single-cell DNA sequencing; scRNA-seq, single-cell RNA sequencing; TCGA-LIHC, The Cancer Genome Atlas Liver Hepatocellular Carcinoma Cohort; TNBC, triple-negative breast cancer; WGD, whole genome doubling.

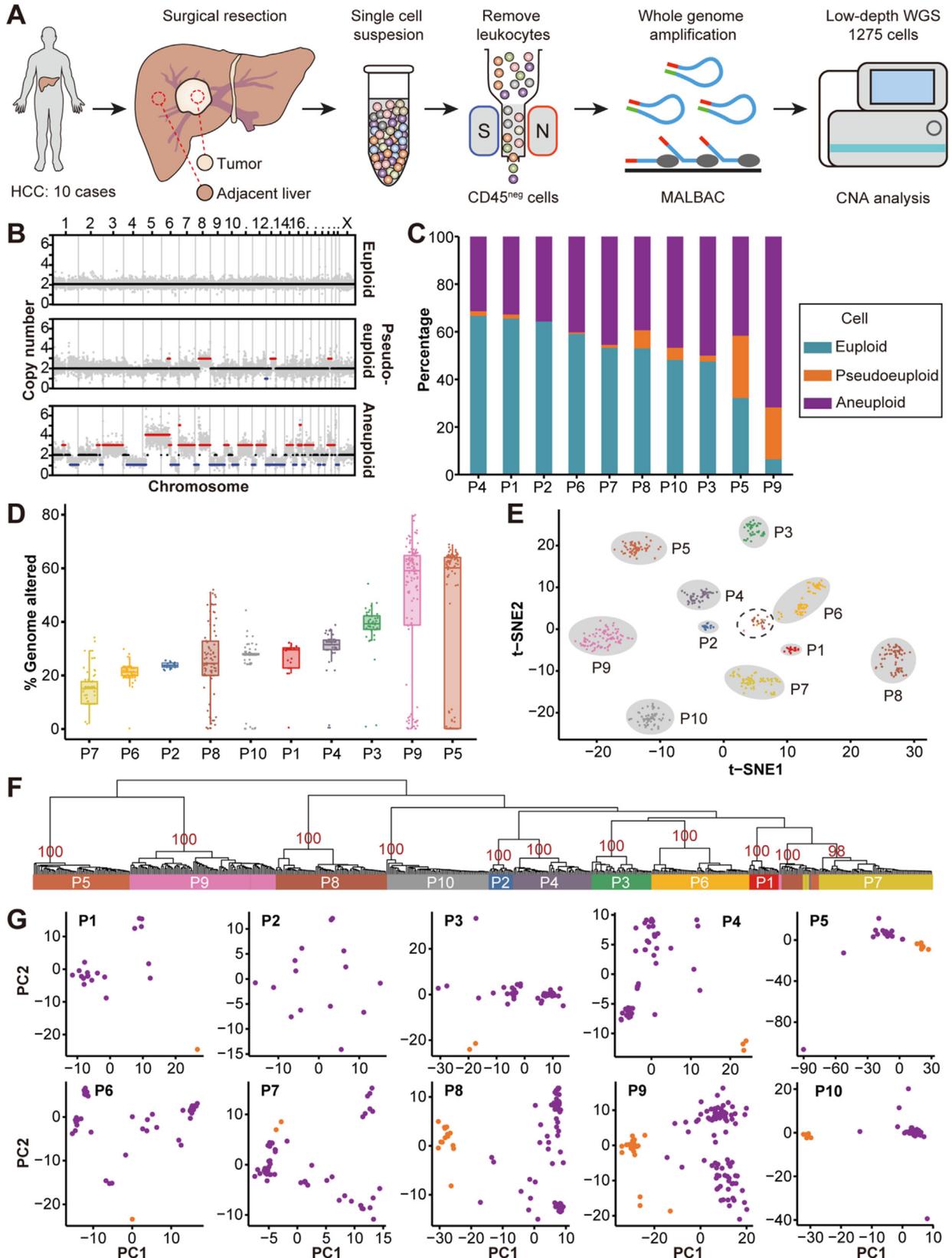
 Most current article

© 2022 by the AGA Institute  
0016-5085/\$36.00

<https://doi.org/10.1053/j.gastro.2021.08.052>

any purification or enrichment step, all cells were selected randomly from the single-cell suspension from each patient after removing leukocytes. Multiple annealing- and

looping-based amplification cycles, strengthened in determining CNAs, were adopted for whole genome amplification.<sup>23-25</sup> Low-depth whole genome sequencing, a



general approach to analyzing CNAs,<sup>24–26</sup> was performed on these single cells (average depth, 0.4×; [Supplementary Table 2](#)). Copy number profiles were calculated from sequence read depth at 500-kb resolution, with the cells from AL as reference ([Supplementary Figures 1 and 2](#)). The pseudo-bulk CNA profiles of each patient did mask many subclonal CNAs affecting numerous drive genes among those single cells. This result may explain the low efficacy of targeted therapies in HCC,<sup>6–9</sup> and highlighted the necessity of our study.

Three cell subpopulations were identified in the tumor tissue ([Figure 1B and C](#), [Supplementary Figure 3](#), and [Supplementary Table 3](#)). Cells without any CNAs were designated as euploid cells. The percentage of euploid cells ranged from 5.7% to 66.5% (median, 52.8%). Presumably, these euploid cells were mainly nonparenchymal cells, like endothelial cells and fibroblasts. Notably, some cells with flat euploid copy number profiles also exhibited approximately 1–10 additional CNAs, defined statistically in the [Supplementary Methods](#). These cells were designated as pseudoeuploid cells and identified in 9 of 10 patients (range, 0%–26.4%; median, 2.2%). The third cell subpopulation, showing profound CNA profiles, was designated as aneuploid cells (range, 31.6%–72.4%; median, 41.2%). The purity of tumor cells across these HCC cases, roughly estimated by the percentage of noneuploid cells consisting of pseudoeuploid and aneuploid cells, ranged from 33.5% to 94.4% (median, 47.2%; mean, 50.91%). The existence of 3 major cell subpopulations in these patients is consistent with the cellular composition observed in TNBC,<sup>22</sup> revealing the variable extent of cellular heterogeneity in HCC.

### Intra- and Intertumor Heterogeneity at Single-Cell Resolution

To explore the heterogeneity among these cells, we focused on noneuploid cells. The percentage of altered genome for each cell, including amplifications and deletions, varied substantially within patients (range of standard deviation, 1.18%–30.16%), as well as among patients (range of median, 15.15%–60.12%) ([Supplementary Table 4](#)), indicating variable extent of intra- and intertumor heterogeneity, respectively ([Figure 1D](#)). The t-distributed stochastic neighbor embedding analysis showed that aneuploid cells from the same patient fell into the same cluster, and cells from different patients fell into different clusters ([Figure 1E](#)). Pseudoeuploid cells from different patients fell into a separate cluster, possibly due to fewer CNA events.

Pairwise Canberra distances among all of the noneuploid cells also resulted in 10 patient-specific clusters of aneuploid cells and 1 cluster of pseudoeuploid cells ([Figure 1F](#)). Similar intra- and intertumor heterogeneity were identified when the same analyses were performed on aneuploid cells only ([Supplementary Figure 4](#)). Collectively, these results show that the extent of intertumor heterogeneity is higher than that of intratumor heterogeneity ([Supplementary Figure 5](#)), consistent with our previous observation in bulk tumors.<sup>6,23,27</sup>

A previous report of TNBC found that tumor cells from the same patient all fell into compact clusters by principal component analysis, indicating stable clonal expansion of tumor cells and limited ITH.<sup>22</sup> Interestingly, aneuploid cells from our studies exhibited a rather dispersed distribution pattern compared with that shown in TNBC, and could not group into 1–3 compact clusters, for instance, P7, P8, and P9 ([Figure 1G and Supplementary Figure 4D](#)). Apparently, HCC tumors have more diverse clonal compositions than TNBC, suggesting a high extent of ITH unlikely to be explained by the PCNE model.

### Diverse Clonal Composition in P9 Indicated a Dual-Phase Copy Number Evolution Model

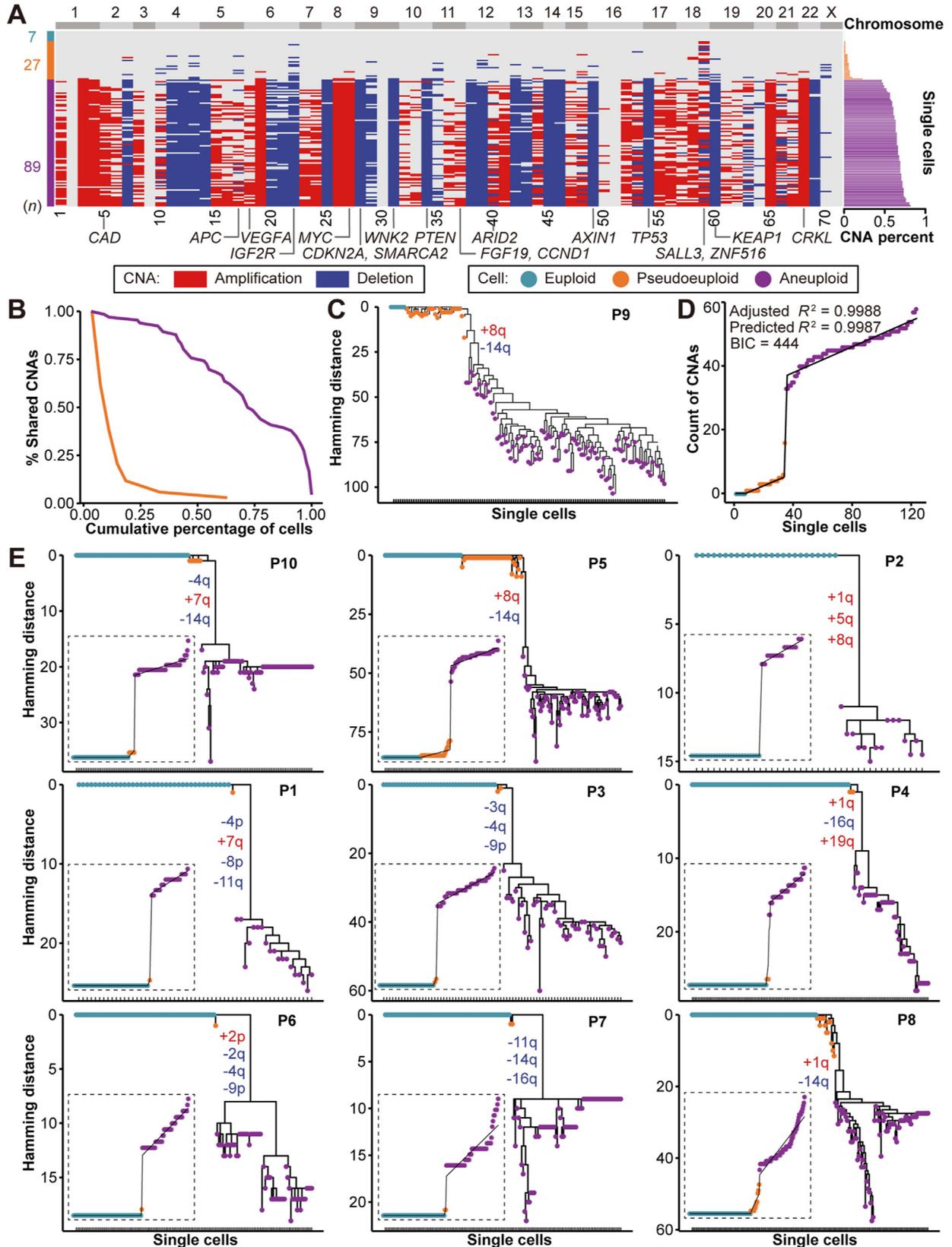
We next explored the evolution model behind the high ITH of HCC tumors. P9 was first selected as an example. In P9, we collected 123 cells, comprising 7 euploid, 27 pseudoeuploid, and 89 aneuploid cells ([Figure 2A](#)). To quantitatively assess CNA profiles of each cell, we calculated a trinary event matrix to treat all CNA events equally for downstream analysis based on the original CNA profiles in [Supplementary Figure 1 \(Supplementary Figure 2\)](#).<sup>22</sup> We identified a total of 66 CNAs across the whole genome binned into 71 regions in P9 ([Figure 2A and Supplementary Figure 6](#)). Pseudoeuploid cells harbored a range of 1 to 16 CNAs (mean, 3.52). For aneuploid cells, a range of 33 to 58 CNAs were identified (mean, 46.07). Cumulative shared percentage of CNAs showed that aneuploid cells shared a much higher percentage of CNAs than pseudoeuploid cells ([Figure 2B](#)). This comparison implies that clonal diversity is large at a relatively early stage, and then diminishes along with tumor progression, indicating potential clonal expansion.

Then we examined the evolution of CNA heterogeneity by constructing a maximum-parsimony phylogenetic tree based on the CNA events of all the single cells ([Figure 2C](#)). Notably, this tree exhibited a short trunk and highly diversified branches, showing extensive and gradual

**Figure 1.** scDNA-seq revealed the CNA heterogeneity landscape in 10 HCC cases. (A) Workflow of scDNA-seq of 10 HCC cases. MALBAC, multiple annealing and looping-based amplification cycle; WGS, whole genome sequencing. (B) Copy number profiles of representative cells from 3 cell subpopulations in the tumor tissue. Red and blue denote amplification and deletion, respectively. (C) Proportions of 3 cell subpopulations across 10 patients. (D) Boxplots for the percentage of altered genome among noneuploid cells from each patient. Cells are colored according to patient. (E) The t-distributed stochastic neighbor embedding (t-SNE) plot of all noneuploid cells. Gray shades denote the clusters of aneuploid cells. Dotted line denotes the cluster of pseudoeuploid cells. (F) Hierarchical clustering of noneuploid cells. Approximately Unbiased (AU) *P* values calculated by pvclust are labeled for each cluster. (G) Principal component analysis (PCA) of the CNA profiles of noneuploid cells.

accumulation of CNAs as cells progressed from euploid to aneuploid genomes. When CNA counts were plotted along single cells, a clear and sharp rise of CNA counts from

pseudoeuploid cells (n = 16) to aneuploid cells (n = 33) was observed, indicating a short burst of CNAs at a relatively early stage before the most recent common ancestor



(Figure 2D).<sup>28,29</sup> This short CNA burst was consistent with that of the PCNE model reported in TNBC,<sup>22</sup> during which the short CNA burst was followed by a rather stable clonal expansion with no more than 5 additional CNA events, arguing that selection force played only a marginal role after the disastrous genomic alterations. However, the divergence among aneuploid cells in P9 reached up to 25, with a median of 13 (Figure 2C and Supplementary Table 5), indicating a significant gradual phase of CNA accumulation. Therefore, we concluded that both PCNE and gradual copy number evolution (GCNE) coexisted in P9, and the marked gradual phase contributed significantly to the higher extent of ITH of P9.

To quantitatively assess the evolution model in P9, the following 3 mathematical fittings were compared: a linear fitting for the GCNE model, a one-step fitting for the PCNE model, and a combined fitting for a novel DPCNE model (Figure 2D). A total of 4 metrics were combined to evaluate 3 models, including adjusted  $R^2$ , predicted  $R^2$ , Bayesian information criterion, and Akaike information criterion. Our DPCNE model (adjusted  $R^2 = 0.9988$ , predicted  $R^2 = 0.9987$ ) outperformed both GCNE model (adjusted  $R^2 = 0.9417$ , predicted  $R^2 = 0.9414$ ,  $P < 2.22E-16$ ) and PCNE model (adjusted  $R^2 = 0.9848$ , predicted  $R^2 = 0.9846$ ,  $P < 2.22E-16$ , ANOVA test). Therefore, CNAs in P9 followed the DPCNE model, that is, a punctuated burst of CNAs followed by a marked gradual phase, which in turn shaped the high extent of ITH in P9.

### Dual-Phase Copy Number Evolution Was Identified Across 10 Patients, Exhibiting Variable Extent of Gradual Evolution Phase

To explore whether DPCNE fitted other patients, we constructed maximum-parsimony phylogenetic trees for all patients (Figure 2E). Trees for most patients exhibited short trunks and long branches, indicating dominant gradual CNA accumulation. Then we looked into aneuploid cells in these patients (P1 and P2 not included because fewer than 20 aneuploid cells were captured), the divergence of CNAs ranged from 8 to 25 (mean, 15.5; median, 13), much higher than the 1–3 proposed in the PCNE model,<sup>22</sup> indicating that tumor cells continually acquired CNAs during the late stage of tumor progression (Supplementary Table 5). For instance, in P8, the CNAs in the gradual phase almost doubled those in the punctuated phase.

Notably, DPCNE performed significantly better than the other 2 models across all patients, suggesting that DPCNE is a better model to describe the clonal evolution pattern in

HCC tumors (Supplementary Figure 7 and Supplementary Table 6). Variable proportions of CNAs in different phases of fitting curves among the 10 patients imply that PCNE and GCNE are not mutually exclusive. On the contrary, both evolutionary patterns contribute to shaping the clonal evolution of HCC tumors.

### Prolonged Gradual Phase of Dual-Phase Copy Number Evolution Correlated With Higher Cellular Heterogeneity and Early Tumor Recurrence

In P9, P6, P7, and P8, we observed a clear advantage of DPCNE fitting over the other 2 fittings, consistent with the prolonged gradual evolution phase in these 4 patients (Figure 3A). This group of patients was designated as the G-group. In the rest of the patients, the fitting results of DPCNE showed a relatively small but significant advantage over those of the PCNE model, consistent with relatively short gradual evolution phases. This group of patients was designated as the P-group. Quantitative analysis showed that all of the patients followed a DPCNE model exhibiting variable extent of gradual phase.

Next, we examined the ITH at single-cell resolution by calculating the pairwise distances of CNA profiles (PWD-CNA) between aneuploid cells in each patient (Figure 3B and Supplementary Table 7). Interestingly, the order of patients by median PWD-CNA value was similar to that by model fitting shown in Figure 3A. We quantitatively compared the PWD-CNA of each tumor with other tumors (Figure 3C and Supplementary Table 8). The most significant differences were observed between G-group and P-group tumors, and comparisons within the same group showed less significant values. Then we pooled the PWD-CNA values of each group together and found that the PWD-CNA of the G-group was significantly higher than that of the P-group (Figure 3D;  $P < 2.22E-16$ , Wilcoxon rank sum test). These results indicate that our classification of P-group and G-group tumors based on model fitting is robust and the G-group tumors with a prolonged gradual phase manifest a higher extent of ITH.

To explore the clinical relevance of our DPCNE model, we investigated whether P-group and G-group tumors were associated with clinical outcomes. We defined cell-ITH as the median value of PWD-CNA of a certain tumor to represent the ITH at the single-cell level. As expected, G-group tumors showed significant higher cell-ITH values than the P-group tumors (Figure 3E;  $P = .001$ , one-tailed  $t$  test). Of note, patients in the G-group had a median disease-free

**Figure 2.** Dual-phase copy number evolution observed in 10 patients with HCC. (A) Trinary event matrices for P9 (gray: 0, neutral; red: 1, amplification; blue: -1, deletion). Color bands to the left show the cell subpopulations, green, euploid cells, orange, pseudoeuploid cells, purple, aneuploid cells. Number of cells are labeled. CNA event IDs are labeled below the heatmap. Potential driver genes affected by the CNAs are labeled accordingly. The percentage of CNA events detected in each single cell are labeled to the right. (B) Cumulative curve of shared CNAs by the number of pseudoeuploid and aneuploid cells, respectively. (C) Maximum-parsimony tree rooted by euploid cells for P9. (D) Model fitting of CNAs in P9. Adjusted  $R^2$ , predicted  $R^2$ , and Bayesian information criterion (BIC) are displayed. (E) Maximum-parsimony trees for other 9 patients. Model fitting results of each patient are embedded.



### Integrative Analysis Identified *CAD* as a Potential Driver Gene for the G-Group

To explore potential driver genes for the prolonged gradual phase in the G-group, we performed genome-wide comparison of CNAs between P-group and G-group tumors (Figure 3G, Supplementary Table 9). Many well-known driver genes located in CNAs shared by both groups, for instance, *TP53* and *AXIN1*. Several driver CNAs were enriched in the G-group, for instance, amplifications in *ARID2*, and losses of *TSC1*<sup>30</sup> and *WNK2*.<sup>31</sup> The enrichment of those genes associated with poor prognosis in the G-group supports our classification of P-group and G-group tumors based on evolutionary patterns and, more importantly, corroborates our finding that G-group tumors are associated with early tumor recurrence.

We designed an integrative strategy based on the 2 features of G-group tumors, shorter DFS and higher ITH, to screen the genes in the G-group enriched CNAs (Figure 3H). We first used the bulk RNA-seq data of 17 patients with HCC with distinct DFS (BCLC B stage, 9 cases that we reported previously<sup>32</sup>). These patients were divided into the early-recurrence group (7 cases, DFS < 6 months)<sup>33</sup> and the late-recurrence group (10 cases, DFS > 24 months, Supplementary Table 1). A total of 1146 differentially expressed genes were identified, including 613 genes up-regulated and 533 genes down-regulated in the early-recurrence group (Supplementary Table 10). Then we identified 19 amplified and 28 deleted genes in the G-group that exhibited concordant expression status with the bulk RNA-seq data (Figure 3H and I and Supplementary Table 11).

Next, we systematically assessed the associations between 47 candidate genes with ITH in the largest published HCC dataset—The Cancer Genome Atlas Liver Hepatocellular Carcinoma Cohort (TCGA-LIHC).<sup>4</sup> We applied a well-accepted strategy of counting subclones to infer the extent of ITH for these single-region tumors (Supplementary Table 12).<sup>34</sup> Notably, a novel candidate gene, *CAD*, exhibited the highest positive correlation with the extent of ITH ( $R = 0.267$ ,  $P = 6.44E-06$ , Pearson correlation) (Figure 3J, Supplementary Figure 8, and Supplementary Table 13). The expression of *CAD* was significantly higher in the “ITH-high” group in comparison with the “ITH-low” group (Figure 3K;  $P = .0092$ , Wilcoxon rank sum test). *CAD*, encoding the carbamoyl-phosphate synthetase 2, aspartate transcarbamylase and dihydroorotase (*CAD*), is a key enzyme of pyrimidine synthesis.<sup>35</sup> Altogether, our integrative analysis identified *CAD* correlates with higher ITH and early tumor recurrence.

### Immunohistochemical Staining Analysis Validated That *CAD* Protein Expression Was Associated With Early Tumor Recurrence

We next examined the *CAD* expression among 1196 liver tumors, comprising 4 published bulk RNA-seq datasets. Higher *CAD* expression correlated with shorter DFS and overall survival (OS) in both the TCGA-LIHC database (Figure 4A; 364 cases, DFS,  $P = 2.1E-05$ ; OS,  $P = .00034$ ) and the KMplotter-HCC database (Figure 4B; 364 cases, DFS,  $P = .0036$ ; OS,  $P = .00015$ , log rank test). In addition, higher *CAD* expression was associated with shorter OS (DFS not available) in both our primary liver cancer cohort (Figure 4C; 309 cases,  $P = 9.2E-06$ )<sup>23</sup> and Gao et al's<sup>36</sup> cohort (Figure 4D; 159 cases,  $P = 6.4E-05$ , log rank test). Meanwhile, *CAD* expression was significantly higher in the tumor compared with AL tissues in both the TCGA-LIHC dataset (Figure 4E; 419 cases,  $P < 2.22E-16$ ) and the Gao et al's cohort (Figure 4F; 318 cases,  $P < 2.22E-16$ , Wilcoxon rank sum test).

To further validate the prognostic value of *CAD*, we performed immunohistochemistry (IHC) of *CAD* on an additional set of 202 HCC cases (Figure 4G). Consistently, *CAD* protein expression in tumors was higher than that in AL tissues (Figure 4H;  $P = .0083$ , Wilcoxon rank sum test). We then divided the HCC cases into early-recurrence and late-recurrence groups and found that early-recurrence tumors did show higher *CAD* expression (Figure 4I;  $P = 5.4E-05$ , Wilcoxon rank sum test). Then we divided these HCC cases into 2 groups by the mean *CAD* expression. We found that *CAD* high tumors did show significant shorter DFS ( $P = .00026$ , log rank test) and OS (Figure 4J;  $P = .0055$ , log rank test). Collectively, high *CAD* expression is a promising biomarker for high extent of ITH and early recurrence of HCC. These results also show the robustness of our screening strategy based on the DPCNE model, demonstrating that the entire approach can be used to identify candidate genes that may be exploited as biomarkers or therapies.

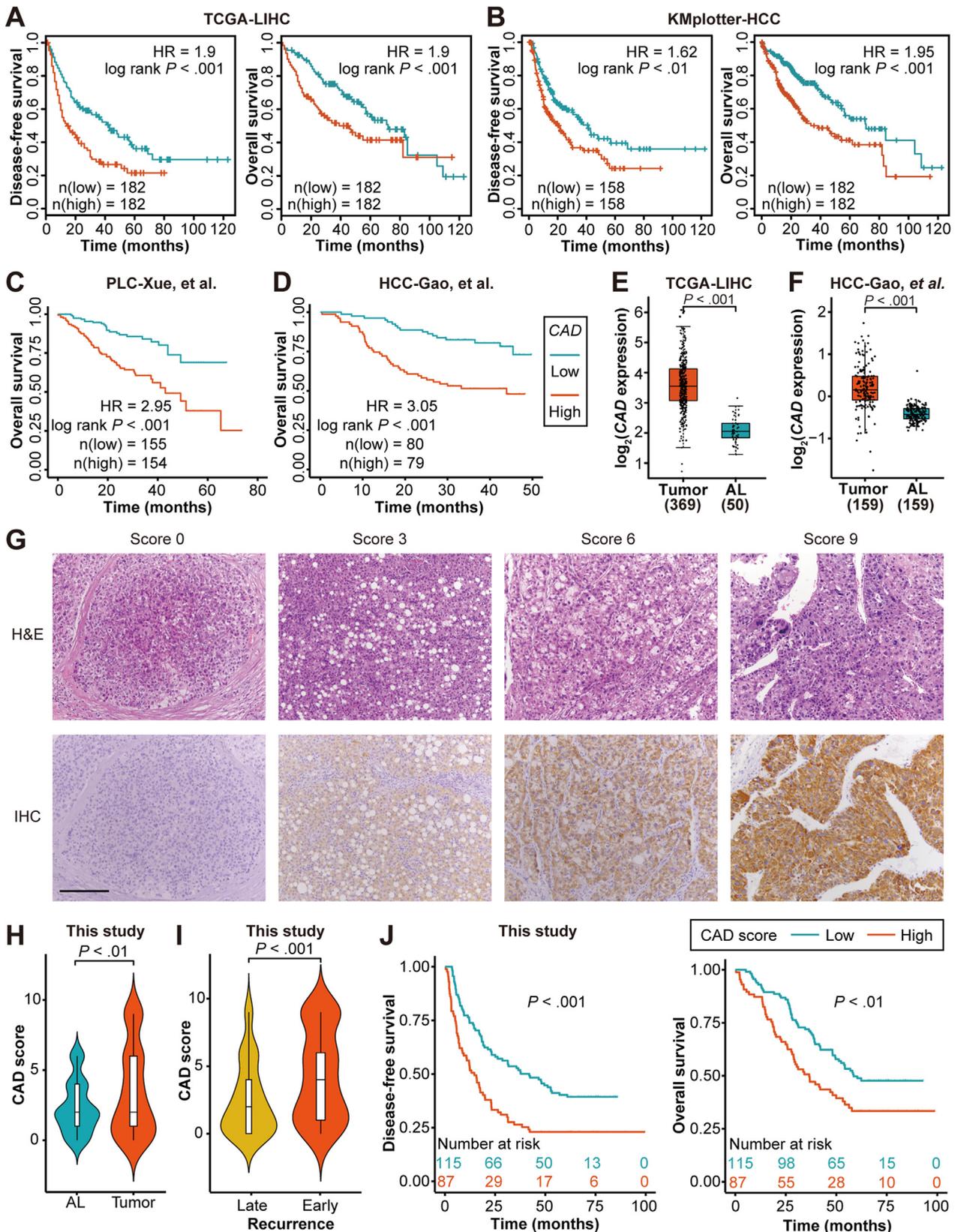
### Dual-Phase Copy Number Evolution Model Validated by Single-Cell Bulk RNA Sequencing Data of Hepatocellular Carcinoma and Published Single-Cell DNA-Sequencing Datasets of Other Cancer Types

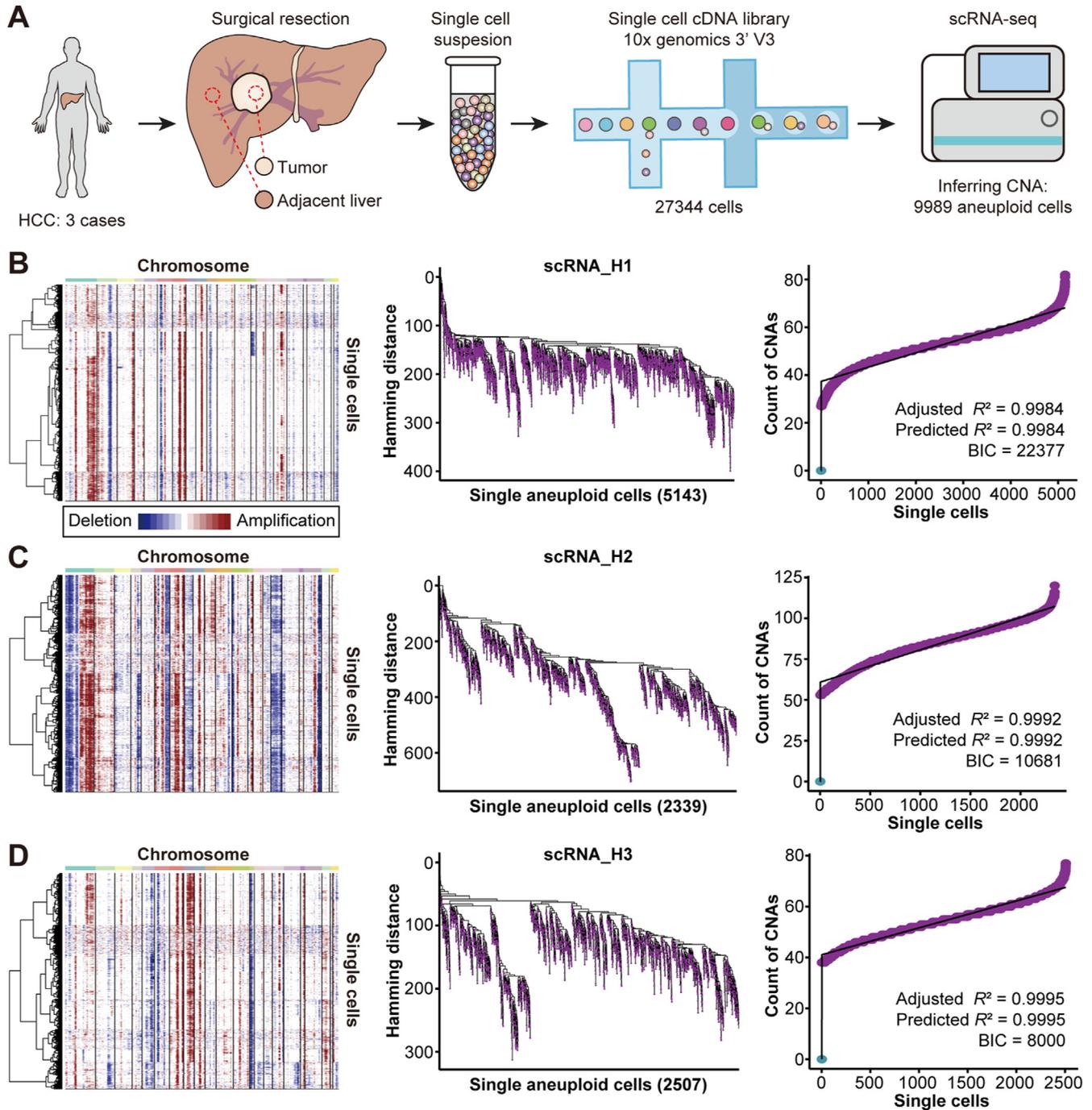
One limitation of scDNA-seq is the low throughput. We further validated our DPCNE model by the high-throughput

**Figure 3.** Classification of P-group and G-group and identification of *CAD*. (A) Model fitting results of each patient with gradual, punctuated, and dual-phase model, respectively. Adjusted  $R^2$  and Bayesian information criterion (BIC) are displayed. (B) Distribution of PWD-CNA values among aneuploid cells for each patient. (C) Pairwise comparisons of PWD-CNA values among 10 patients, Wilcoxon rank sum test. (D) Comparison of pooled PWD-CNA values between P-group and G-group patients, Wilcoxon rank sum test. (E) Comparison of cell-ITH between P-group and G-group patients, one-tailed  $t$  test. (F) Comparison of DFS between P-group and G-group patients, log rank test. (G) Comparison of CNA events and selected genes between P-group and G-group patients. Orange, G-group enriched genes, green, P-group enriched genes, black, shared genes. (H) Screening of potential driver genes for the G-group tumors. (I) Volcano plot of differentially expressed genes comparing early-recurrence to late-recurrence tumors. (J) Correlation analysis between *CAD* expression and the extent of ITH from the TCGA-LIHC cohort, Pearson correlation. (K) Comparison of *CAD* expression between ITH-low and ITH-high groups from the TCGA-LIHC cohort, Wilcoxon rank sum test.

scRNA-seq method. Briefly, we used a droplet-based scRNA-seq sequencing platform (10X Genomics) to portray the cellular landscape of 3 HCC tumors, designated as scRNA\_H1,

scRNA\_H2, scRNA\_H3 (Figure 5A).<sup>27</sup> Among the 27,344 cells harvested, we discriminated 9989 aneuploid/malignant cells and 4380 euploid/nonmalignant cells by inferring large-scale





**Figure 5.** DPCNE model validated by scRNA-seq data of 3 HCC tumors. (A) Workflow of scRNA-seq in 3 HCC cases. (B–D) CNA profiles (left), maximum-parsimony trees (middle), and model fitting results of CNA count (right) for scRNA\_H1 (B), scRNA\_H2 (C) and scRNA\_H3 (D). Red and blue denote amplification and deletion, respectively. Adjusted  $R^2$ , predicted  $R^2$ , and Bayesian information criterion (BIC) are displayed.

**Figure 4.** High expression level of *CAD* was associated with poor prognosis in HCC. (A, B) DFS and OS of HCC tumors from the TCGA-LIHC database (A) and the KMplotter-HCC database (B) stratified into *CAD*-high and *CAD*-low groups, respectively, log rank test. (C, D) OS of 309 primary liver cancer patients from Xue et al<sup>23</sup> (C) and 159 HCC patients from Gao et al<sup>36</sup> (D) stratified into *CAD*-high and *CAD*-low groups, respectively, log rank test. (E) Comparison of *CAD* expression between tumor and AL tissues from the TCGA-LIHC database, Wilcoxon rank sum test. (F) Comparison of *CAD* expression between tumor and AL tissues from Gao et al, Wilcoxon rank sum test. (G) IHC staining of *CAD* in our validation cohort of 202 HCC cases. Representative images of staining scores are shown. Scale bar: 200  $\mu$ m. (H) Violin plots of *CAD* expression scores in the tumor and AL tissues of 202 HCC cases, Wilcoxon rank sum test. (I) Violin plots of *CAD* expression of 202 cases stratified into early-recurrence and late-recurrence groups, Wilcoxon rank sum test. (J) Kaplan-Meier plots for DFS and OS for 202 HCC cases separated by the average *CAD* expression, log rank test.

chromosomal CNAs (Figure 5B–D, Supplementary Table 14). We noticed that the CNA profiles inferred from scRNA-seq data were noisier than those from scDNA-seq, lacking power in unambiguously determining small CNAs (Figure 5B–D). As a result, we were not able to identify pseudoeuploid cells in these samples. This result further supports the necessity of scDNA-seq in investigating HCC clonal evolution.

Nevertheless, the quality of the data was sufficient to test the DPCNE model. The tumor cells (here referred to as malignant cells) determined from the scRNA-seq data were about 10 times of the tumor cells (here referred to as noneuploid cells) involved in our scDNA-seq data. The CNA profiles of 3 HCC tumors all showed diverse subclonal composition, consistent with our observation from scDNA-seq data (Figure 5B–D). Phylogenetic trees all exhibited short trunks followed by highly diversified branches. CNA count plot of 3 HCC tumors all showed a sharp rise of CNA counts from normal cells ( $n = 0$ ) to aneuploid cells ( $n = 27$ ,  $n = 53$ , and  $n = 38$ , respectively), similar to the short burst of CNAs in the punctuated phase. Then, the cells continually acquired additional CNAs in a gradual manner (from 27/53/38 to 82/120/77, accounting for 67.07%/55.83%/50.65% of all CNAs). Model fitting results showed that DPCNE model outperformed both GCNE ( $P < 2.22E-16$ ) and PCNE ( $P < 2.22E-16$ , ANOVA test) (Supplementary Figure 9 and Supplementary Table 6). These results validated our DPCNE model in a much larger number of cells.

To further explore whether our DPCNE model applies to other cancer types, we re-analyzed 2 published scDNA-seq datasets from TNBC and colorectal cancer.<sup>22,37</sup> Notably, DPCNE model outperformed PCNE and GCNE across all cases in both colorectal cancer and TNBC (Supplementary Figure 10 and Supplementary Table 15). Comparing the fitting results of different models across these cases revealed that 7 of 9 patients with colorectal cancer and 2 of 12 patients with TNBC could be categorized into the G-group, indicating a prolonged gradual phase in these cases. These results suggest that our DPCNE model is robust and applies to other cancer types.

### Ploidy-Resolved Single-Cell DNA Sequencing Showed That Diploid- and Polyploid-Aneuploid Cells Had a Common Clonal Origin

Ploidy heterogeneity is an important facet of liver cellular heterogeneity because hepatocytes can be diploid or polyploid, where polyploid hepatocytes account for 25%–50% of hepatocytes in human livers.<sup>38,39</sup> To explore ploidy heterogeneity in HCC, we performed fluorescence-activated cell sorting (FACS) on the tumor and AL tissues from 10 newly recruited HCC cases (Figure 6A). Because ploidy classification based on DNA content is accurate when hepatocytes are in G0/G1, yet may be confused with cell division when hepatocytes have replicated their DNA in G2/M, we further performed Ki-67 staining to assess the proliferation state of tissues.<sup>40</sup> Very few cells in AL tissues were actively cycling (Supplementary Figure 11 and Supplementary Table 16), indicating that ploidy quantification based on DNA

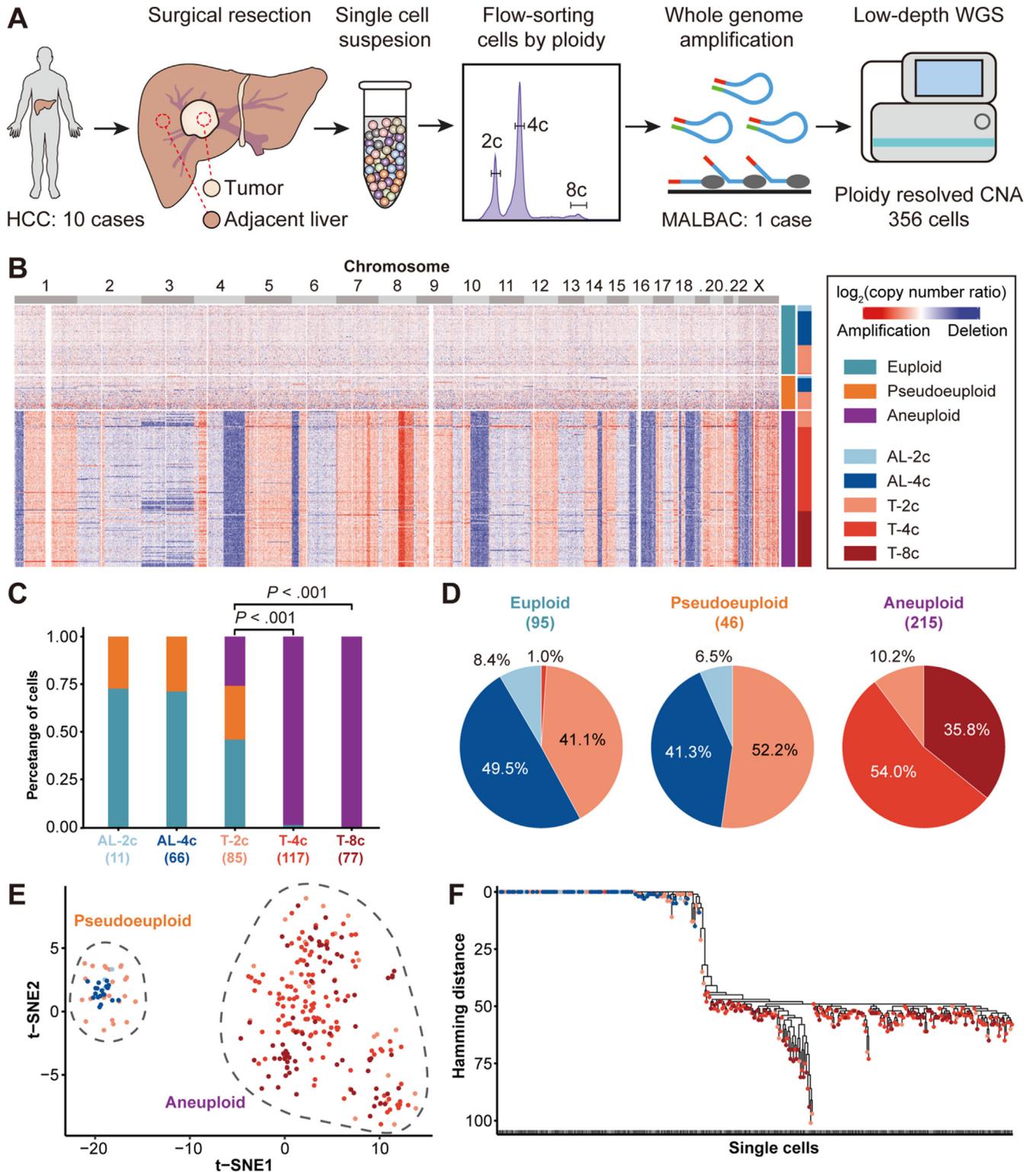
content is accurate for AL tissues. By contrast, variable extent of cycling cells was observed among tumor tissues (range, 3%–40%; median, 10%). To minimize the potential confusion of cell division in ploidy quantification, we performed Ki-67 correction on the original cellular composition from FACS. As a result, most tumors ( $n = 9$  of 10) contained a majority of diploid cells while P20 had an obvious proportion of polyploid tumor cells (diploid, 33.39%, tetraploid, 35.58%, and octaploid, 2.5%) (Supplementary Figures 12 and 13 and Supplementary Table 16). This result is consistent with previous reports that HCCs are predominantly diploid.<sup>41–44</sup>

To further gain ploidy resolution to the cellular ITH of HCC, we combined FACS with scDNA-seq to perform ploidy-scDNA-seq on 1 selected case, P20, which had the most abundant polyploid cells. We successfully performed ploidy-scDNA-seq on 356 cells covering 5 FACS subsets (AL-2c, 11 cells; AL-4c, 66 cells; T-2c, 85 cells; T-4c, 117 cells; and T-8c, 77 cells) (Figure 6B, Supplementary Table 17) and determined the absolute CNA profiles of each subset (Supplementary Figure 14). As expected, cells from the AL were mostly euploid (Figure 6C). Surprisingly, 25.9% ( $n = 22$  of 85) of T-2c cells were aneuploid, standing in sharp contrast to that 99.1% ( $n = 116$  of 117) of T-4c and 100% ( $n = 77$  of 77) of T-8c cells were aneuploid. This result clearly showed that aneuploid cells are highly enriched in the polyploid subsets (Figure 6D). Because only 1 T-4c cell was not aneuploid, a ploidy correction rate of 30% based on Ki-67 would have neglectable changes. Furthermore, the results of T-8c clearly illustrated the enrichment of aneuploid among polyploid cells whether these cells are tetraploid in G2/M or octaploid in G0/G1.

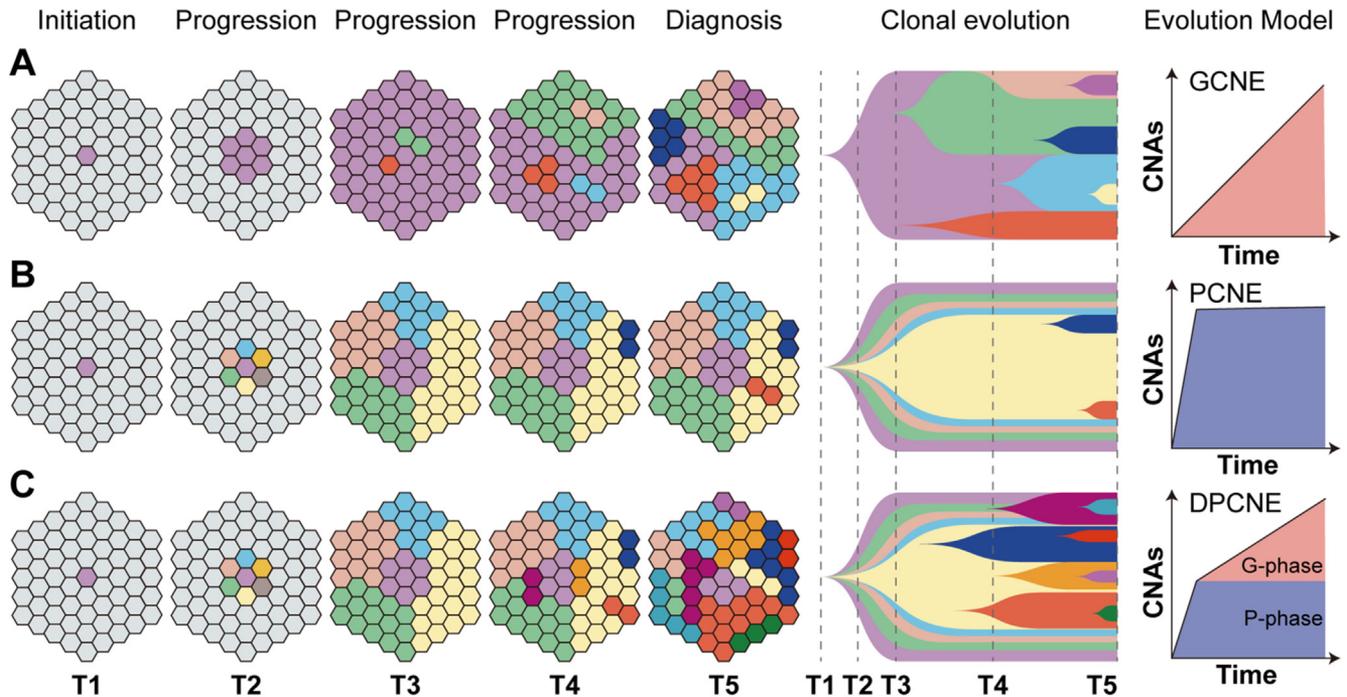
More interestingly, t-distributed stochastic neighbor embedding clustering and phylogenetic tree analysis showed that T-2c, T-4c, and T-8c cells were mixed together rather than clustered separately (Figure 6E and F). This result suggested that diploid- and polyploid-aneuploid cells shared the same clonal origin. An obvious evidence can be seen from subclonal losses of 3p and 3q, which were not enriched in either diploid or polyploid subsets (Figure 6B, Supplementary Figure 15). Instead, they can be found across T-2c, T-4c, and T-8c, suggesting that these subclonal CNA events were inherited during the ploidy alteration. Taking the enrichment of aneuploidy in polyploid subsets into account, the highest possibility here is that those polyploid-aneuploid cells were generated by whole genome doubling (WGD) of diploid-aneuploid cells. Therefore, our result from this case showed that aneuploid cells first originated from the diploid subset and then evolved into polyploid subsets. Collectively, our results from FACS and ploidy-scDNA-seq support the diploid origin of HCC tumors, consistent with studies on rodents that diploid cells are more susceptible to tumorigenesis.<sup>40,45,46</sup>

## Discussion

Our DPCNE model showed that punctuated evolution and gradual evolution are not mutually exclusive (Figure 7). Instead, both evolution patterns can coexist in the same



**Figure 6.** FACS analysis of 10 HCC patients and ploidy-scDNA-seq of P20. (A) Workflow of FACS analysis and ploidy-scRNA-seq. MALBAC, multiple annealing and looping-based amplification cycle; WGS, whole genome sequencing. (B) Copy number profiles of all the cells in P20. Row sidebars denote the 3 cell subpopulations and the 5 ploidy subsets. (C) Proportions of 3 cell subpopulations across the 5 ploidy subsets. (D) Proportions of 5 ploidy subsets across the 3 cell subpopulations across. (E) The t-distributed stochastic neighbor embedding (t-SNE) analysis of the CNA profiles of noneuploid cells in P20. Cells are labeled by the 5 ploidy subsets. (F) Maximum-parsimony tree rooted by euploid cells for P20.



**Figure 7.** Three models of copy number evolution. (A) GCNE. (B) PCNE. (C) DPCNE. For each of 3 models, we use 3 different plots to describe the CNA evolution patterns. Five different timing points (T1, T2, T3, T4, and T5) spanning from tumor initiation to progression and diagnosis are denoted accordingly. Different clones and subclones were labeled by color. *Left: Honeycomb plots* show discrete screenshots of tumor subclonal composition at different timing points. *Middle: Clonal evolution plots* show the continuous process of tumor subclonal composition. *Right: Evolution model plots* show the CNA accumulation along time. *Red and blue regions* denote punctuated phase and gradual phase, respectively.

tumor and may drive hepatocarcinogenesis at different stages. Previous studies supporting the PCNE model borrow evidence from catastrophic CNA events, such as chromothripsis and chromoplexy.<sup>17,21</sup> However, those events only account for a small percentage of all CNAs and are often accompanied by other discrete somatic alterations. The co-occurrence of discrete and catastrophic events in most cancer types actually supports our DPCNE model. Therefore, our model reconciles the conflict between gradual evolution and punctuated evolution, implying that it is not a binary choice and both evolution strategies can be exploited by tumor cells, alternatively or parallelly.

Technically, both scDNA-seq and scRNA-seq methods have pros and cons in detecting CNAs.<sup>47</sup> scDNA-seq can accurately determine CNAs,<sup>18,19</sup> making it a standard tool to investigate the complex tumor subclonal structure. Due to the duplication error and allelic dropout inherent in whole genome amplification of single cells that mostly have only 2 copies of DNA, the throughput of scDNA-seq is often limited to hundreds of cells. Increasing the throughput of scDNA-seq comes at the cost of coverage and accuracy. By contrast, scRNA-seq are more robust because messenger RNA is more abundant than DNA in single cells and has many copies. scRNA-seq can be easily integrated to droplet-based platforms, enabling the high-throughput transcriptomic analysis of thousands of cells in a single assay.<sup>27,48,49</sup> However, scRNA-seq cannot directly determine CNAs. Although CNAs can be mathematically inferred from scRNA-seq data, it performs best at arm-level CNAs to distinguish malignant from nonmalignant cells while lacks

power in determining small CNAs to reconstruct the detailed subclonal structure. Of note, we were not able to identify pseudoeuploid cells from our scRNA-seq data. Therefore, scDNA-seq is still the better choice for investigating the CNA landscape. Accurate and high-throughput scDNA-seq methods are needed in the future. In addition, current scDNA-seq and scRNA-seq methods cannot discriminate diploid and polyploid cells.

Whether HCC is composed mainly of diploid or polyploid cells is hotly debated. Older reports suggest that HCCs are predominantly diploid,<sup>41-43</sup> and this has been validated more recently.<sup>44</sup> Numerous studies have found that diploid hepatocytes are more sensitive to HCC formation compared with polyploid hepatocytes, which partially explains why they drive tumorigenesis.<sup>40,45,46</sup> However, opposite observations have found that a subset of HCC with *TP53* mutations are characterized by polyploidy,<sup>50</sup> which seems to stand against the diploid origin of HCC and deny the tumor suppressive role of polyploid hepatocytes. Despite limited results from a single case, our ploidy-scDNA-seq analysis led to the following key observations: aneuploid cells were enriched in polyploid subsets rather than the diploid subset; and among those aneuploid cells in the tumor, polyploid subsets shared a common origin with their diploid counterparts. The highest possibility here, at least in this single case, is that aneuploid tumor cells originate from the diploid subset and underwent WGD to generate those polyploid subsets, but not vice versa. In this scenario, the reported *TP53*-mutant-enriched polyploid tumors may

originate from diploid hepatocytes, yet exhibited a higher frequency of WGD in comparison to other diploid tumors.<sup>50</sup> In agreement with this, a pan-cancer analysis reveals that *TP53* mutations are highly associated with WGD.<sup>51</sup> Therefore, our results may reconcile the conflict reports of diploid/polyploid nature of HCC tumors and indicate that previous seeming contrary results may not be inconsistent, and all actually support the diploid origin of HCC tumors. Whether all of those polyploid HCC tumors result from WGD of diploid predecessors warrants further large-cohort studies.

At last, we identified a gene, *CAD*, involved in pyrimidine synthesis, as a novel prognostic biomarker for HCC. Previous studies found that *CAD* up-regulation increases tumor proliferation by rewiring urea cycle metabolism in cancer to support anabolism.<sup>35,52</sup> Therefore, urea cycle-related metabolites in the plasma and urine may serve as biomarkers for the G-group patients, higher cell-ITH and early tumor recurrence. Our study sheds light on potential links among altered tumor metabolism, highly diversified clonal substructure, and poor tumor prognosis.

## Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at [www.gastrojournal.org](http://www.gastrojournal.org), and at <http://doi.org/10.1053/j.gastro.2021.08.052>.

## References

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209–249.
- Villanueva A. Hepatocellular carcinoma. *N Engl J Med* 2019;380:1450–1462.
- Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers* 2016;2:16018.
- The Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 2017;169:1327–1341.e23.
- Dhanasekaran R, Nault JC, Roberts LR, et al. Genomic medicine and implications for hepatocellular carcinoma prevention and therapy. *Gastroenterology* 2019;156:492–509.
- Xue R, Li R, Guo H, Guo L, et al. Variable intra-tumor genomic heterogeneity of multiple lesions in patients with hepatocellular carcinoma. *Gastroenterology* 2016;150:998–1008.
- Zhai W, Lim TK, Zhang T, et al. The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. *Nat Commun* 2017;8:4565.
- Torrecilla S, Sia D, Harrington AN, et al. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. *J Hepatol* 2017;67:1222–1231.
- Losic B, Craig AJ, Villacorta-Martin C, et al. Intratumoral heterogeneity and clonal evolution in liver cancer. *Nat Commun* 2020;11:291.
- McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 2017;168:613–628.
- PCAWG Transcriptome Core Group, et al. Genomic basis for RNA alterations in cancer. *Nature* 2020;578:129–136.
- Navin NE. The first five years of single-cell cancer genomics and beyond. *Genome Res* 2015;25:1499–1507.
- Hou Y, Guo H, Cao C, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* 2016;26:304–319.
- Duan M, Hao J, Cui S, et al. Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing. *Cell Res* 2018;28:359–373.
- Ma L, Hernandez MO, Zhao Y, et al. Tumor cell biodiversity drives microenvironmental reprogramming in liver cancer. *Cancer Cell* 2019;36:418–430.e6.
- Graham TA, Sottoriva A. Measuring cancer evolution from the genome. *J Pathol* 2017;241:183–191.
- Turajlic S, Sottoriva A, Graham T, et al. Resolving genetic heterogeneity in cancer. *Nat Rev Genet* 2019;20:404–416.
- Craig AJ, von Felden J, Garcia-Lezana T, et al. Tumour evolution in hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* 2020;17:139–152.
- Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. *Science* 2013;339:1546–1558.
- Sottoriva A, Kang H, Ma Z, et al. A Big Bang model of human colorectal tumor growth. *Nat Genet* 2015;47:209–216.
- Baca SC, Prandi D, Lawrence MS, et al. Punctuated evolution of prostate cancer genomes. *Cell* 2013;153:666–677.
- Gao R, Davis A, McDonald TO, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet* 2016;48:1119–1130.
- Xue R, Chen L, Zhang C, Fujita M, et al. Genomic and transcriptomic profiling of combined hepatocellular and intrahepatic cholangiocarcinoma reveals distinct molecular subtypes. *Cancer Cell* 2019;35:932–947.e8.
- Gao Y, Ni X, Guo H, Su Z, Ba Y, et al. Single-cell sequencing deciphers a convergent evolution of copy number alterations from primary to circulating tumor cells. *Genome Res* 2017;27:1312–1322.
- Zong C, Lu S, Chapman AR, et al. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 2012;338:1622–1626.
- Baslan T, Kendall J, Ward B, et al. Optimizing sparse sequencing of single cells for highly multiplex copy number profiling. *Genome Res* 2015;25:714–724.
- Xiang X, Liu Z, Zhang C, Li Z, et al. IDH mutation subgroup status associates with intratumor heterogeneity and the tumor microenvironment in intrahepatic cholangiocarcinoma. *Adv Sci (Weinh)* 2021;8(17):e2101230.
- Nik-Zainal S, Van Loo P, Wedge DC, et al. The life history of 21 breast cancers. *Cell* 2012;149:994–1007.
- Zhang C, Zhang L, Xu T, Xue R, Yu L, et al. Mapping the spreading routes of lymphatic metastases in human colorectal cancer. *Nat Commun* 2020;11:1993.

30. **Ho DWH, Chan LK**, Chiu YT, et al. TSC1/2 mutations define a molecular subset of HCC with aggressive behaviour and treatment implication. *Gut* 2017;66:1496–1506.
31. **Zhou SL, Zhou ZJ, Hu ZQ, Song CL**, et al. Genomic sequencing identifies WNK2 as a driver in hepatocellular carcinoma and a risk factor for early recurrence. *J Hepatol* 2019;71:1152–1163.
32. **Chen L, Liu D, Yi X**, et al. The novel miR-1269b-regulated protein SVEP1 induces hepatocellular carcinoma proliferation and metastasis likely through the PI3K/Akt pathway. *Cell Death Dis* 2020;11:320.
33. Hoshida Y, Villanueva A, Kobayashi M, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med* 2008;359:1995–2004.
34. Andor N, Graham TA, Jansen M, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* 2016;22:105–113.
35. **Rabinovich S, Adler L, Yizhak K**, et al. Diversion of aspartate in ASS1-deficient tumours fosters de novo pyrimidine synthesis. *Nature* 2015;527:379–383.
36. **Gao Q, Zhu H, Dong L, Shi W, Chen R**, et al. Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell* 2019;179:561–577 e22.
37. **Bian S, Hou Y, Zhou X**, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* 2018;362:1060–1063.
38. **Donne R, Saroul-Ainama M, Cordier P**, et al. Polyploidy in liver development, homeostasis and disease. *Nat Rev Gastroenterol Hepatol* 2020;17:391–405.
39. Wilkinson PD, Duncan AW. Differential roles for diploid and polyploid hepatocytes in acute and chronic liver injury. *Semin Liver Dis* 2021;41:42–49.
40. Lin YH, Zhang S, Zhu M, et al. Mice with increased numbers of polyploid hepatocytes maintain regenerative capacity but develop fewer hepatocellular carcinomas following chronic liver injury. *Gastroenterology* 2020;158:1698–1712.e14.
41. Anti M, Marra G, Rapaccini GL, et al. DNA ploidy pattern in human chronic liver diseases and hepatic nodular lesions. Flow cytometric analysis on echo-guided needle liver biopsy. *Cancer* 1994;73:281–288.
42. Ruà S, Comino A, Fruttero A, et al. Flow cytometric DNA analysis of cirrhotic liver cells in patients with hepatocellular carcinoma can provide a new prognostic factor. *Cancer* 1996;78:1195–1202.
43. Saeter G, Schwarze PE, Nesland JM, et al. Diploid nature of hepatocellular tumours developing from transplanted preneoplastic liver cells. *Br J Cancer* 1989;59:198–205.
44. Sladky VC, Knapp K, Szabo TG, et al. PIDDosome-induced p53-dependent ploidy restriction facilitates hepatocarcinogenesis. *EMBO Rep* 2020;21:e50893.
45. Wilkinson PD, Delgado ER, Alencastro F, et al. The polyploid state restricts hepatocyte proliferation and liver regeneration in mice. *Hepatology* 2019;69:1242–1258.
46. Zhang S, Zhou K, Luo X, et al. The polyploid state plays a tumor-suppressive role in the liver. *Dev Cell* 2018;44:447–459.e5.
47. Xue R, Li R, Bai F. Single cell sequencing: technique, application, and future development. *Sci Bull* 2015;60:33–42.
48. Ramachandran P, Matchett KP, Dobie R, et al. Single-cell technologies in hepatology: new insights into liver biology and disease pathogenesis. *Nat Rev Gastroenterol Hepatol* 2020;17:457–472.
49. Dong X, Xue R. Capture a live-photo of tumor infiltrating T cell landscape at single cell resolution. *Cancer Biol Med* 2019;16:619–622.
50. **Bou-Nader M, Caruso S, Donne R**, et al. Polyploidy spectrum: a new marker in HCC classification. *Gut* 2020;69:355–364.
51. Bielski CM, Zehir A, Penson AV, et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet* 2018;50:1189–1195.
52. Keshet R, Szlosarek P, Carracedo A, et al. Rewiring urea cycle metabolism in cancer to support anabolism. *Nat Rev Cancer* 2018;18:634–645.

Author names in bold designate shared co-first authorship.

Received December 11, 2020. Accepted August 25, 2021.

#### Correspondence

Address correspondence to: Ning Zhang, PhD, Translational Cancer Research Center, Peking University First Hospital, No. 8, Xishiku Street, Xicheng District, Beijing 100034, China. e-mail: [zhangning@bjmu.edu.cn](mailto:zhangning@bjmu.edu.cn), or Ruidong Xue, PhD, Translational Cancer Research Center, Peking University First Hospital, No. 8, Xishiku Street, Xicheng District, Beijing 100034, China. e-mail: [rxue@hsc.pku.edu.cn](mailto:rxue@hsc.pku.edu.cn).

#### Acknowledgments

Sequence data are deposited in the Genome Sequence Archive in Beijing Institute of Genomics (<https://bigd.big.ac.cn/gsa>), Chinese Academy of Sciences, under accession numbers HRA000094 and HRA000872. The data deposited and made public are compliant with the regulations of Ministry of Science and Technology of China. Lin Guo, Xianfu Yi, Lu Chen, and Ti Zhang contributed equally to this work.

#### CRedit Authorship Contributions

Lin Guo, PhD (Conceptualization: Lead; Data curation: Lead; Methodology: Lead; Writing – original draft: Lead; Writing – review & editing: Equal). Xianfu Yi, PhD (Conceptualization: Equal; Formal analysis: Lead; Software: Lead; Visualization: Lead; Writing – review & editing: Equal). Lu Chen, MD (Resources: Lead; Validation: Equal). Ti Zhang, MD (Resources: Equal; Validation: Equal). Hua Guo, PhD (Resources: Supporting). Ziyi Chen, PhD (Validation: Supporting). Jinghui Cheng, PhD (Formal analysis: Supporting). Qi Cao, PhD (Software: Supporting). Hengkang Liu, PhD (Formal analysis: Supporting). Chunyu Hou, PhD (Software: Supporting). Lisha Qi, MD (Methodology: Supporting). Zhiyan Zhu, PhD (Methodology: Supporting). Yucun Liu, MD (Resources: Supporting). Ruirui Kong, PhD (Methodology: Supporting). Chong Zhang, PhD (Formal analysis: Supporting). Xiaohua Zhou, PhD (Methodology: Supporting). Zemin Zhang, PhD (Conceptualization: Supporting; Writing – review & editing: Supporting). Tianqiang Song, MD (Resources: Supporting). Ruidong Xue, PhD (Conceptualization: Lead; Formal analysis: Lead; Methodology: Lead; Supervision: Lead; Validation: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead). Ning Zhang, PhD (Conceptualization: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Supervision: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead).

#### Conflicts of interest

This author disclosed the following: Zemin Zhang is a founder of Analytical BioSciences and a consultant for InnoCare Pharma and ArsenalBio. The remaining authors disclose no conflicts.

#### Funding

This work is jointly supported by the National Natural Science Foundation of China (82173035, 81802813, 82030079, 81972656, 81988101, and 81902401), the National Science and Technology Major Project of China (2018ZX10723204), the Michigan Medicine and Peking University Health Science Center Joint Institute for Translational and Clinical Research (BMU2020J1005), Natural Science Foundation of Tianjin (19JCQNJC09000), and China Scholarship Council (201906940003). The authors also received grant support from Sino-Russian Math Center and BioMap Inc.