

Supplementary Information

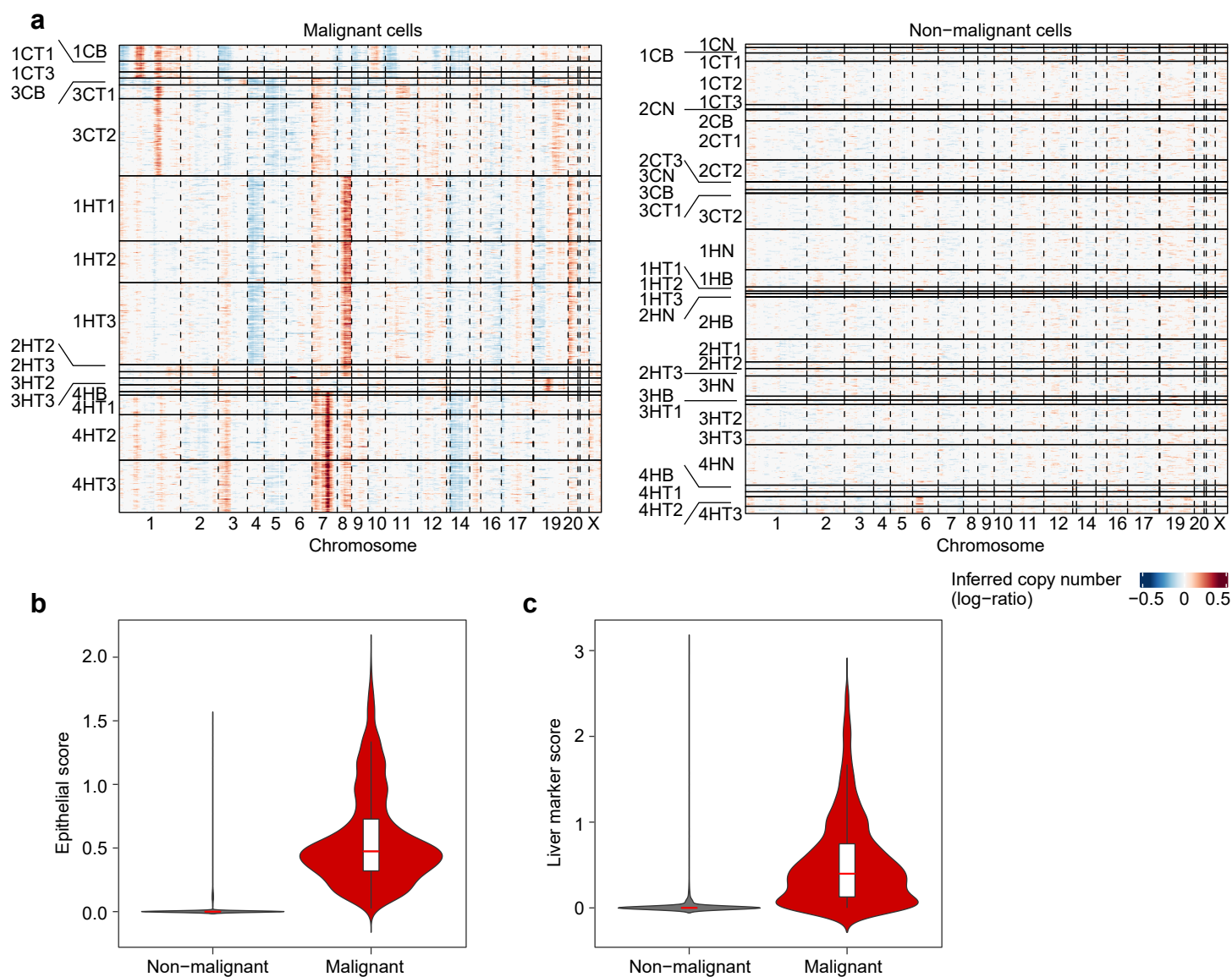
Multiregional single-cell dissection of tumor and immune cells reveals stable lock-and-key features in liver cancer

Lichun Ma, Sophia Heinrich, Limin Wang, Friederike L. Keggenhoff, Subreen Khatib, Marshonna Forgues, Michael Kelly, Stephen M. Hewitt, Areeba Saif, Jonathan M. Hernandez, Donna Mabry, Roman Kloeckner, Tim F. Greten, Jittiporn Chaisaingmongkol, Mathuros Ruchirawat, Jens U. Marquardt, Xin Wei Wang

This file contains:

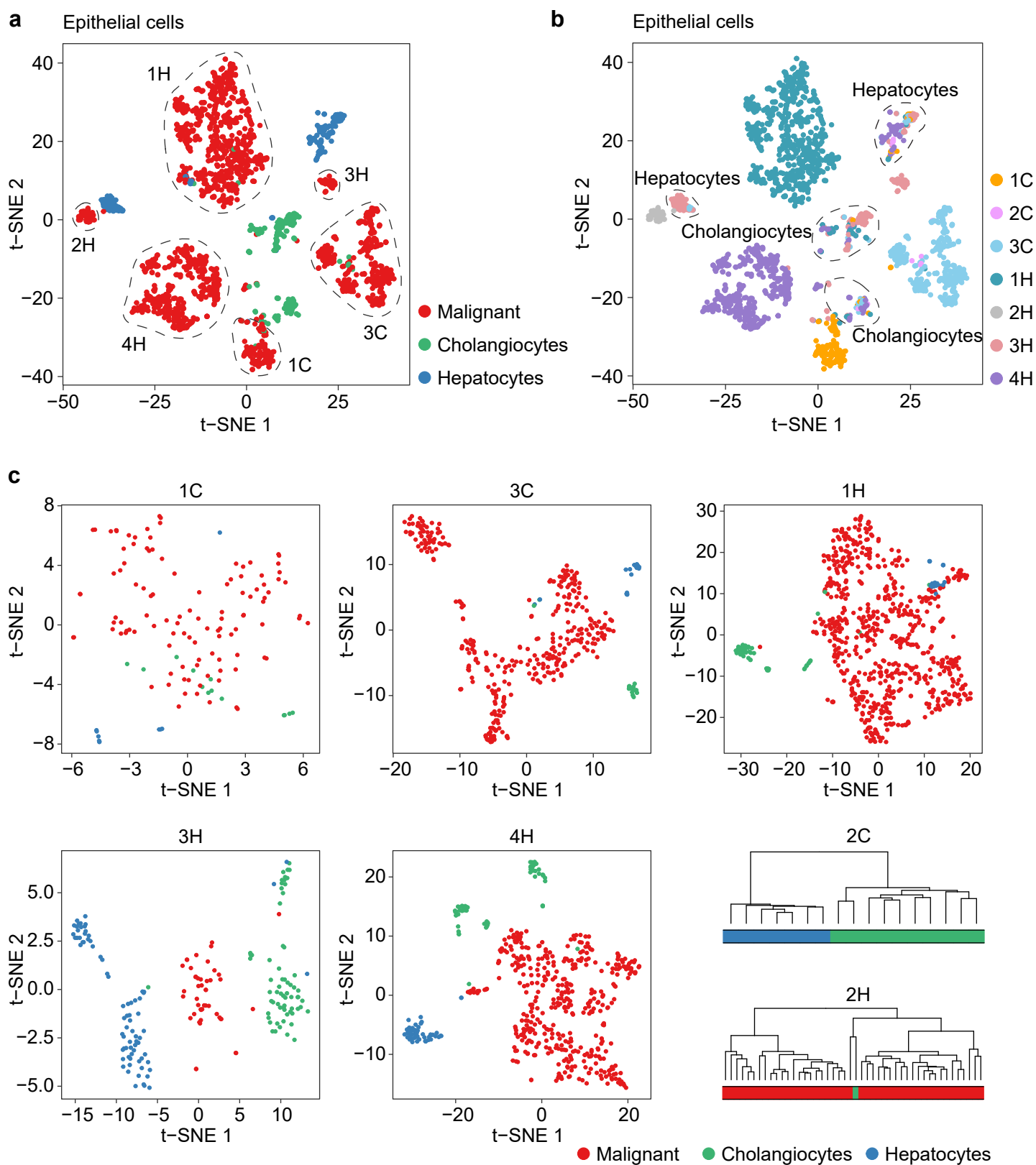
Supplementary Figures 1-17

Supplementary Table 1



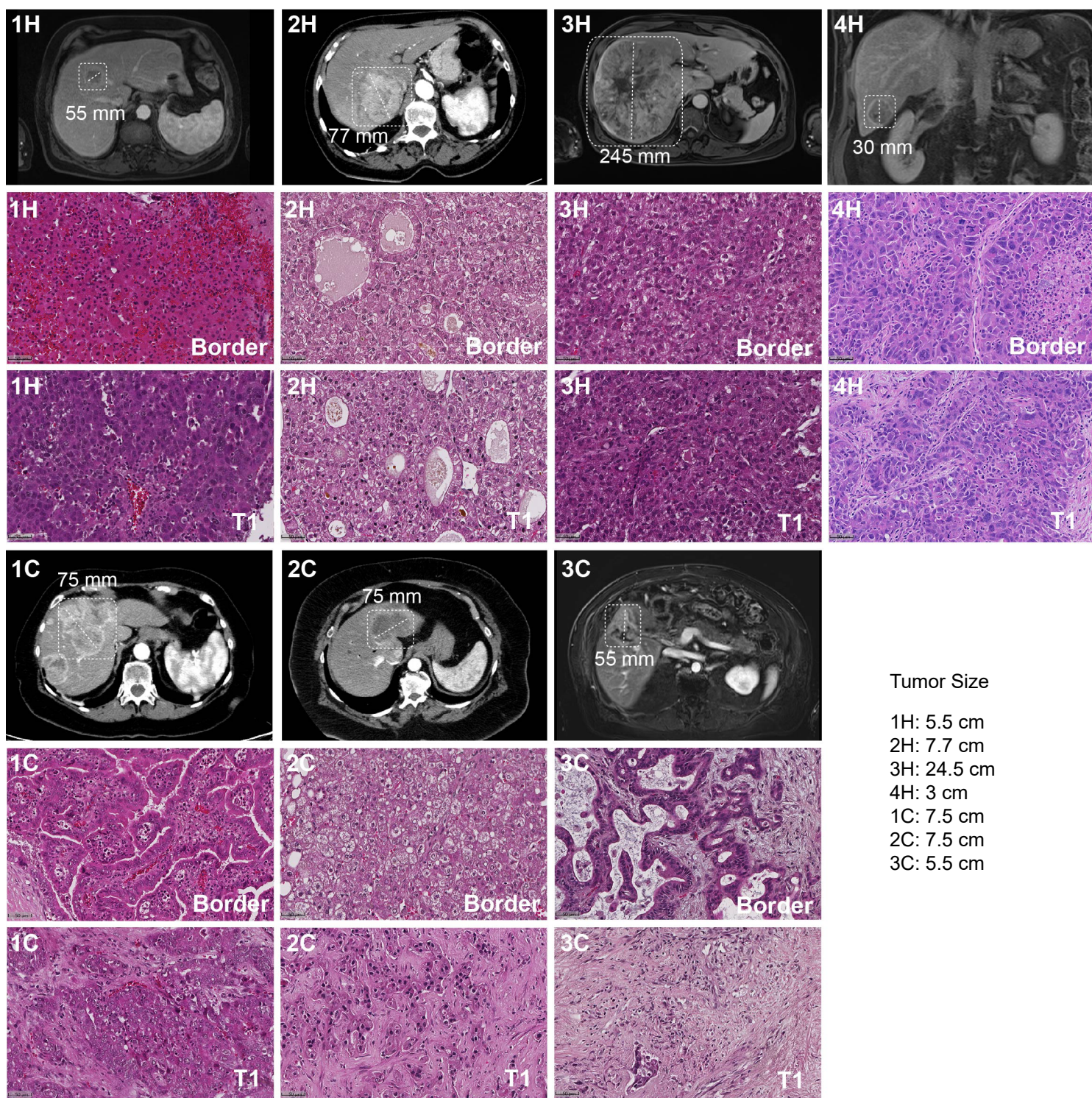
Supplementary Figure 1. Separation of malignant and non-malignant cells.

(a) Inferred large-scale CNVs of malignant cells (left) and non-malignant cells (right). Red, amplifications; blue, deletions. (b and c) Epithelial marker score (b) and liver marker score (c) of malignant cells (n=1,598) and non-malignant cells (n=77,670). Reference cells from normal regions were not included in the violin plot. In the boxplots, the central rectangles span the first quartile to the third quartile, with the segments inside the rectangle corresponding to the median. Whiskers extend 1.5 times the interquartile range.



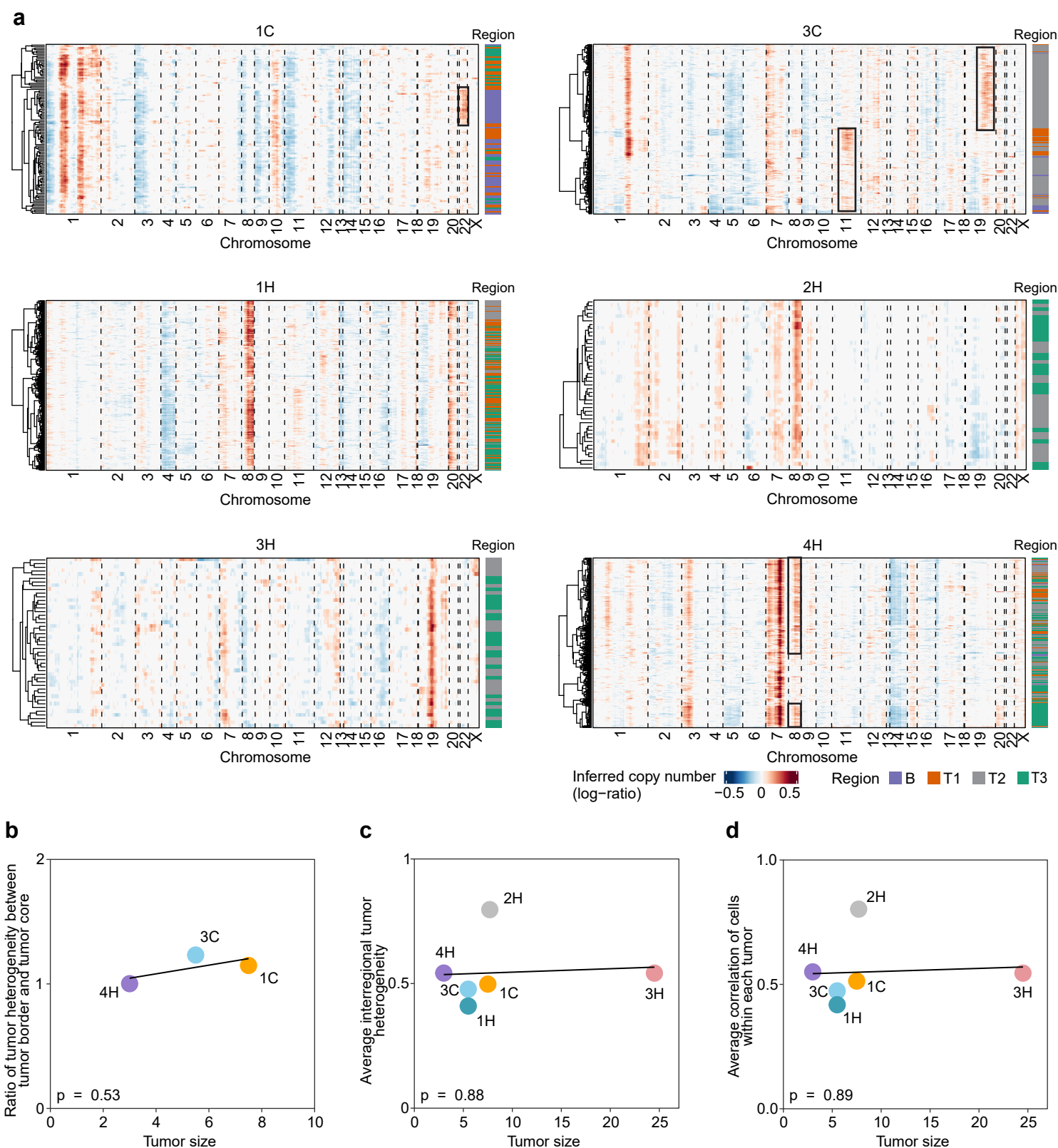
Supplementary Figure 2. Analysis of epithelial cells.

(a and b) t-SNE of epithelial cells from tumor and adjacent normal liver tissues colored by cell types (a) and cases (b). (c) t-SNE or hierarchical clustering of epithelial cells in (a) for each individual case. Due to small number of cells in 2C and 2H, hierarchical clustering was applied due to failure in generating t-SNE plots.



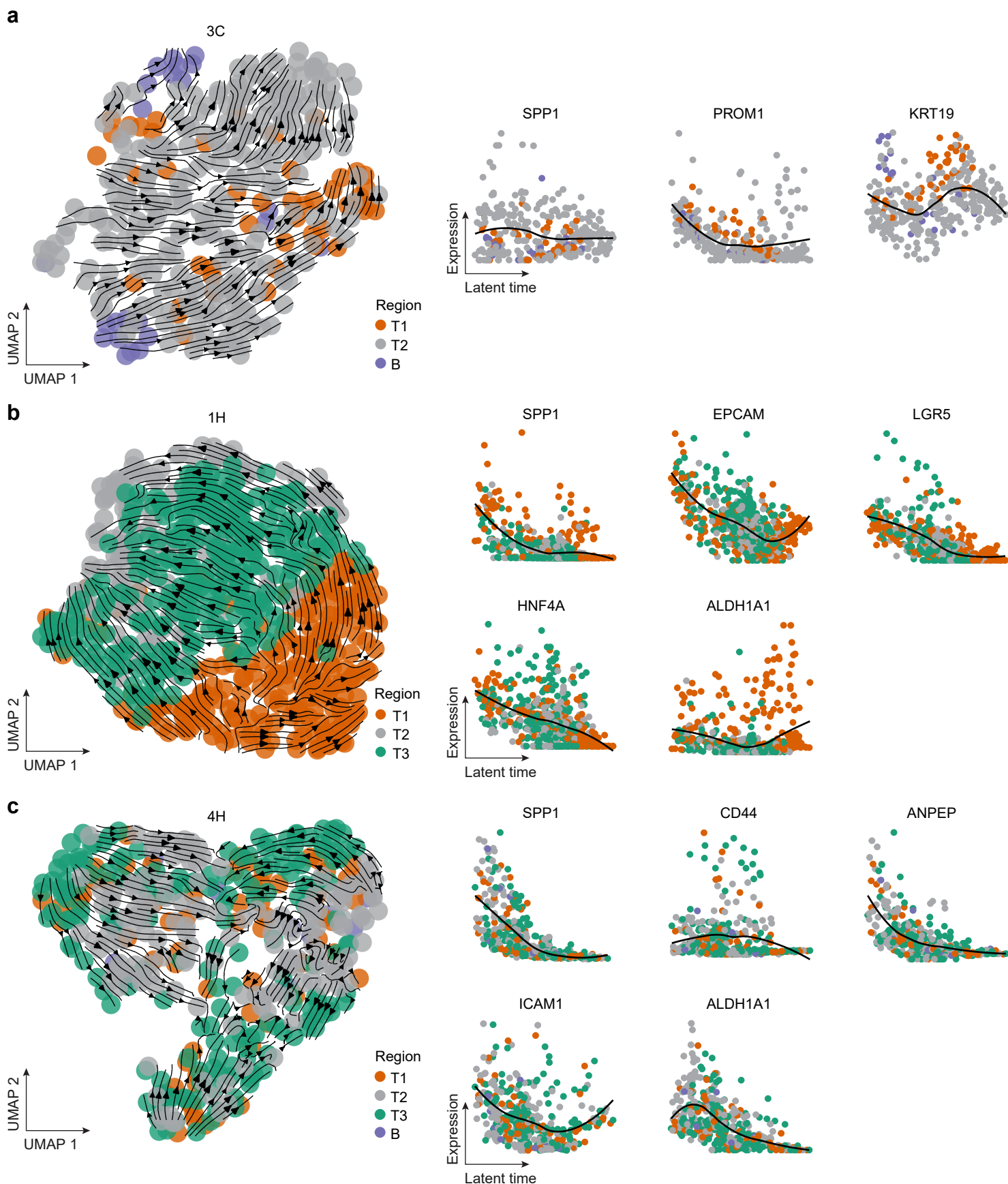
Supplementary Figure 3. Magnetic resonance imaging and histopathology.

Magnetic resonance imaging of all the HCC and iCCA cases and histopathology of tumors collected from tumor border and T1 region. Scale bars, 50 μ m.



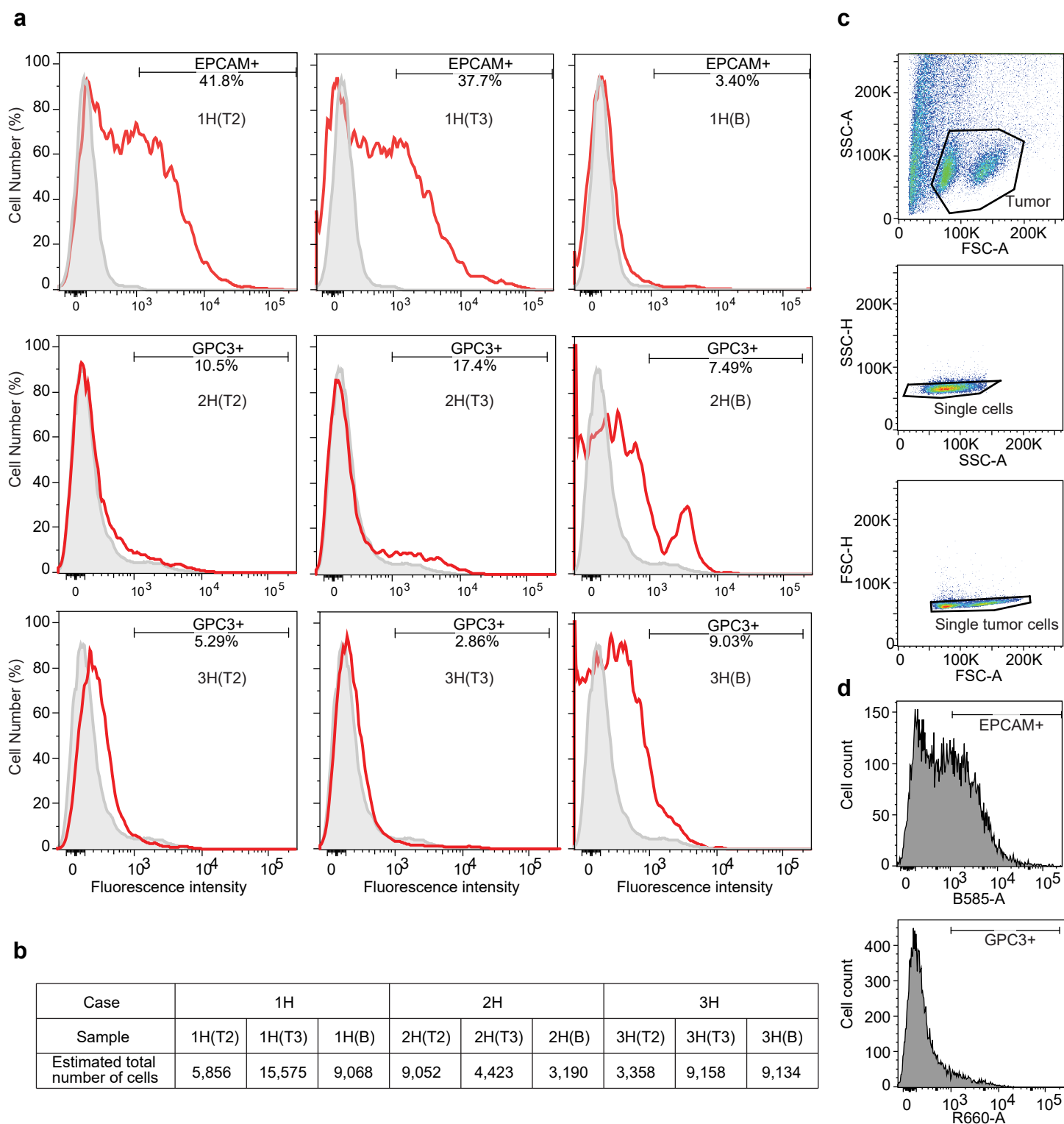
Supplementary Figure 4. Heterogeneity of malignant cells.

(a) Hierarchical clustering of inferred CNVs of malignant cells for each individual case. Examples of clones with distinct CNVs were highlighted using black boxes. (b) The relationship of tumor size and the ratio of tumor heterogeneity (average pair-wise correlation value) between tumor border and tumor core. (c) The relationship of tumor size and the average interregional tumor heterogeneity of each case. (d) The relationship of tumor size and the average pair-wise correlation of all malignant cells within each case. Pearson's correlation test (two-sided) was used in (b-d) to indicate the statistical significance of the relationship. See also Fig. 1f.



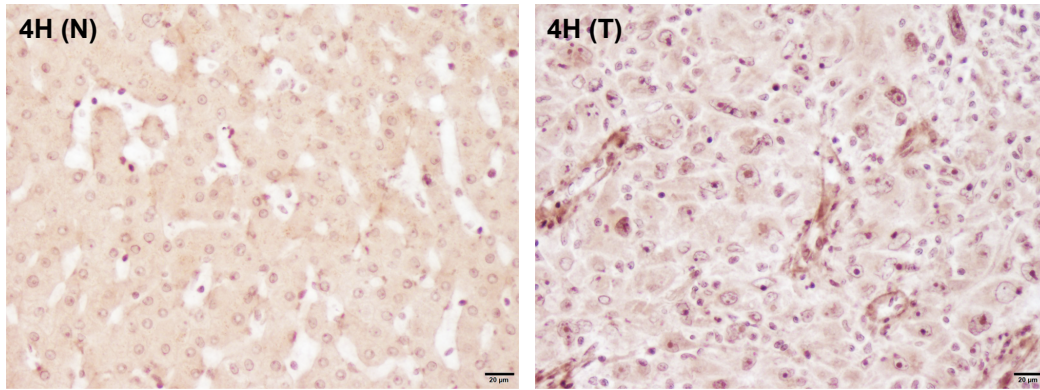
Supplementary Figure 5. Multiregional tumor cell trajectory.

(a-c) RNA velocity of malignant cells from multiple tumor regions and expression of several marker genes (cell stemness-related marker genes and SPP1) along cellular latent time of case 3C (a), 1H (b) and 4H (c).



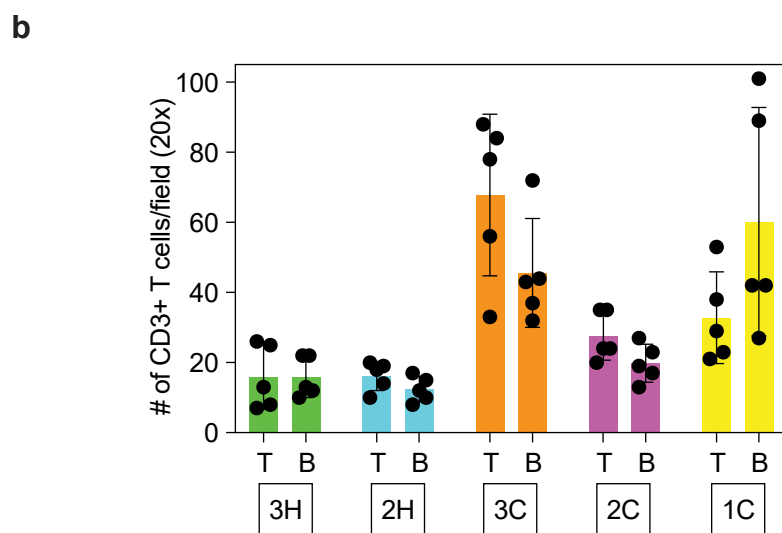
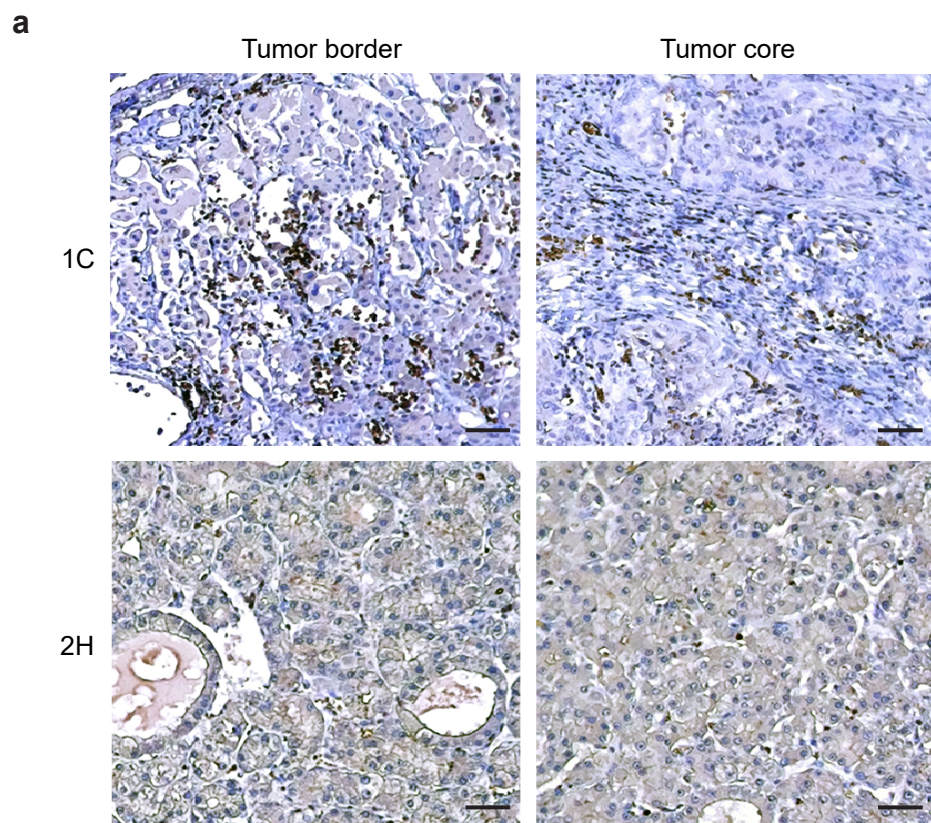
Supplementary Figure 6. FACS of EPCAM+ or GPC3+ cells.

(a) Distribution of EPCAM+ cells from different tumor regions of 1H (first row) and GPC3+ cells from multiple regions of 2H and 3H (second and third rows). The percentage of EPCAM+ or GPC3+ cells was indicated on figures. (b) Estimated total number of cells from the samples in (a). (c and d) Cell gating strategy: cells were sequentially gated (c, from top to bottom) with indicated scatters to obtain a uniform population of single tumor cells by excluding debris, dead cells and clusters, and then gated with cells stained using antibody isotype control to get EPCAM and GPC3 signal as shown in representative samples (d).



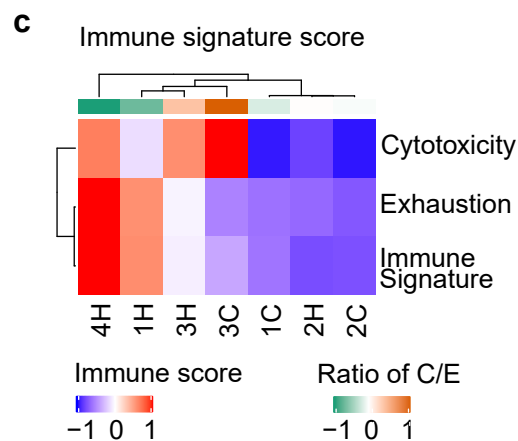
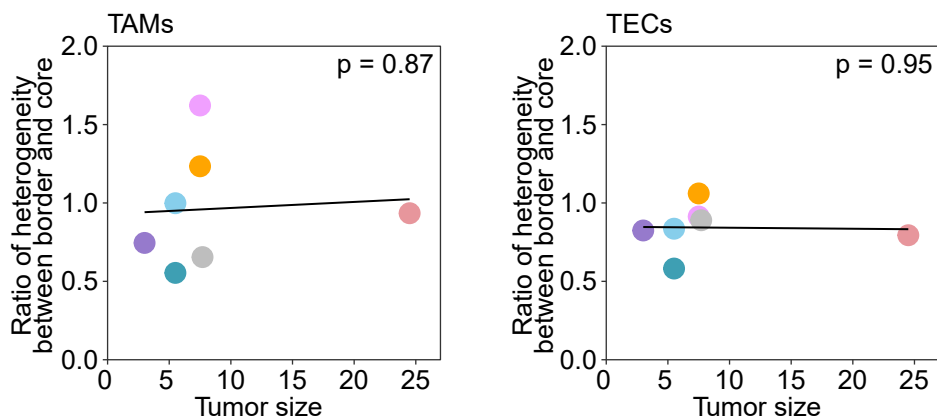
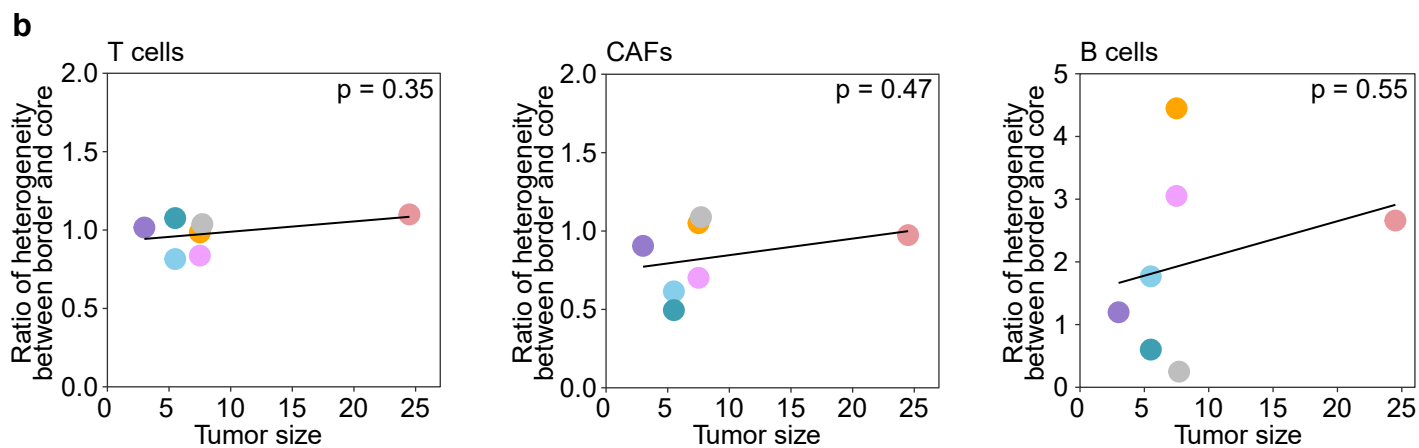
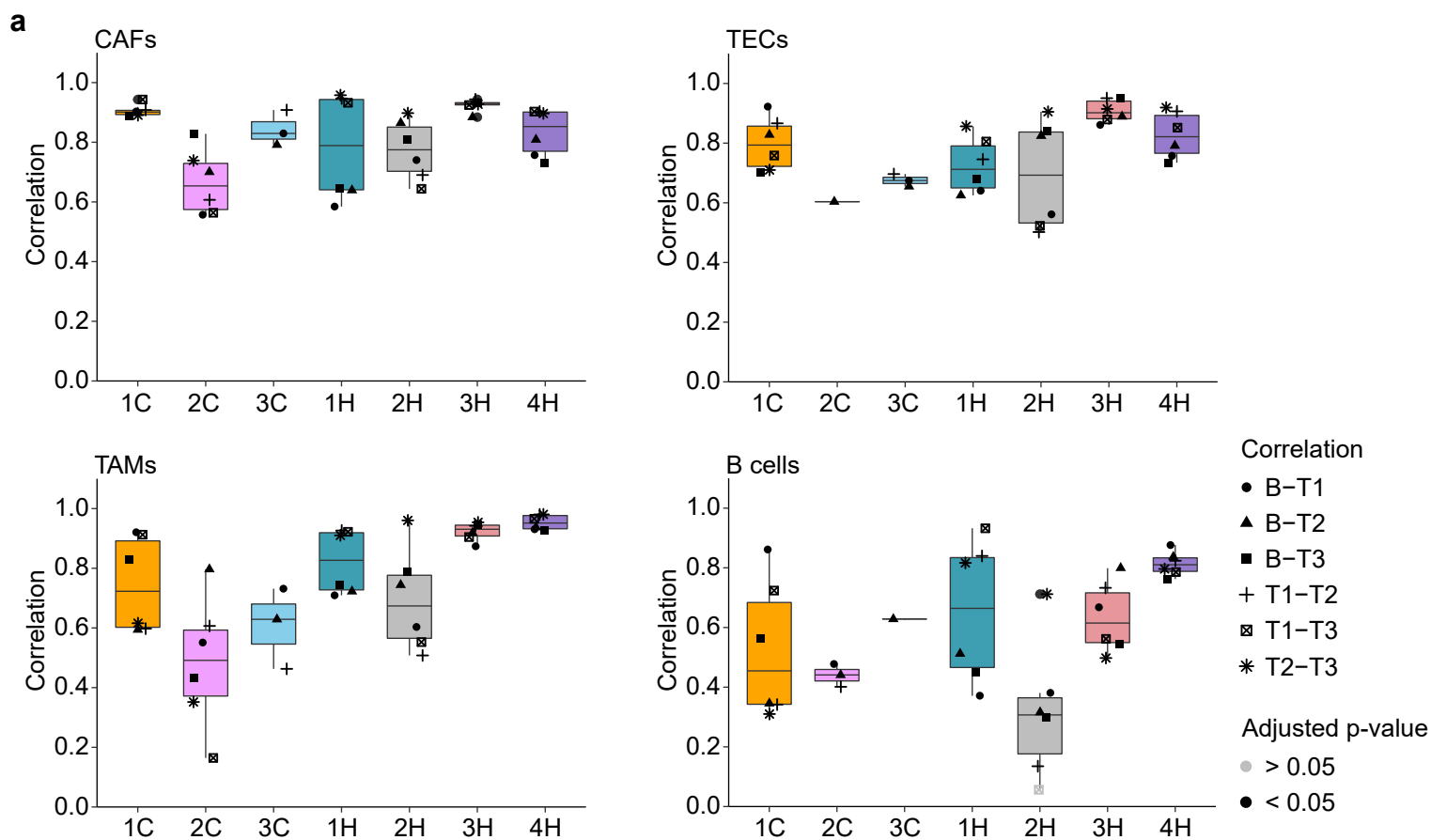
Supplementary Figure 7. Immunohistochemistry of SPP1.

Immunohistochemistry of SPP1 in non-tumor (left) and tumor (right) tissues from case 4H. Scale bars, 20 μm.



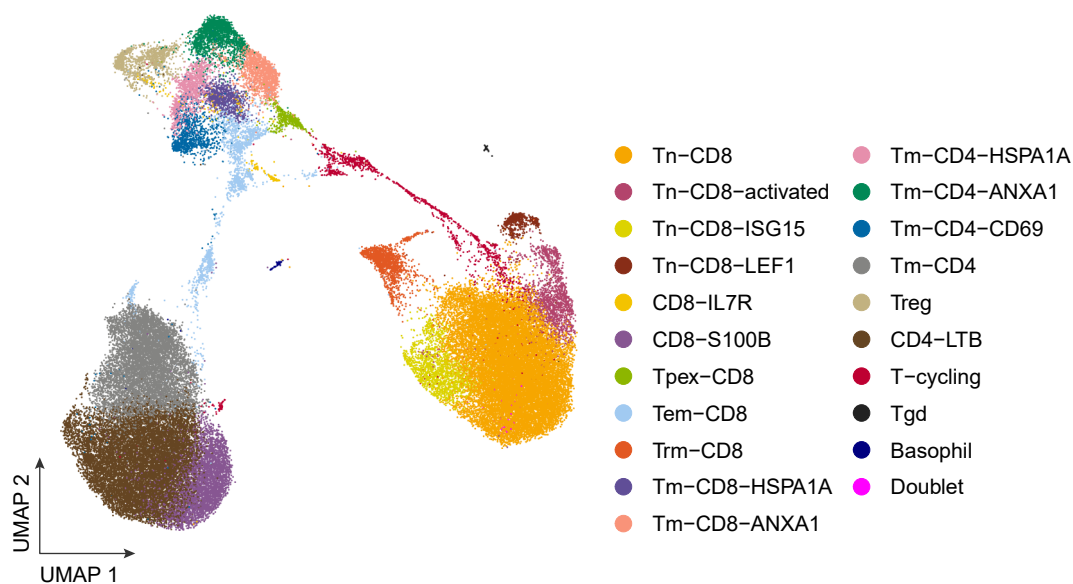
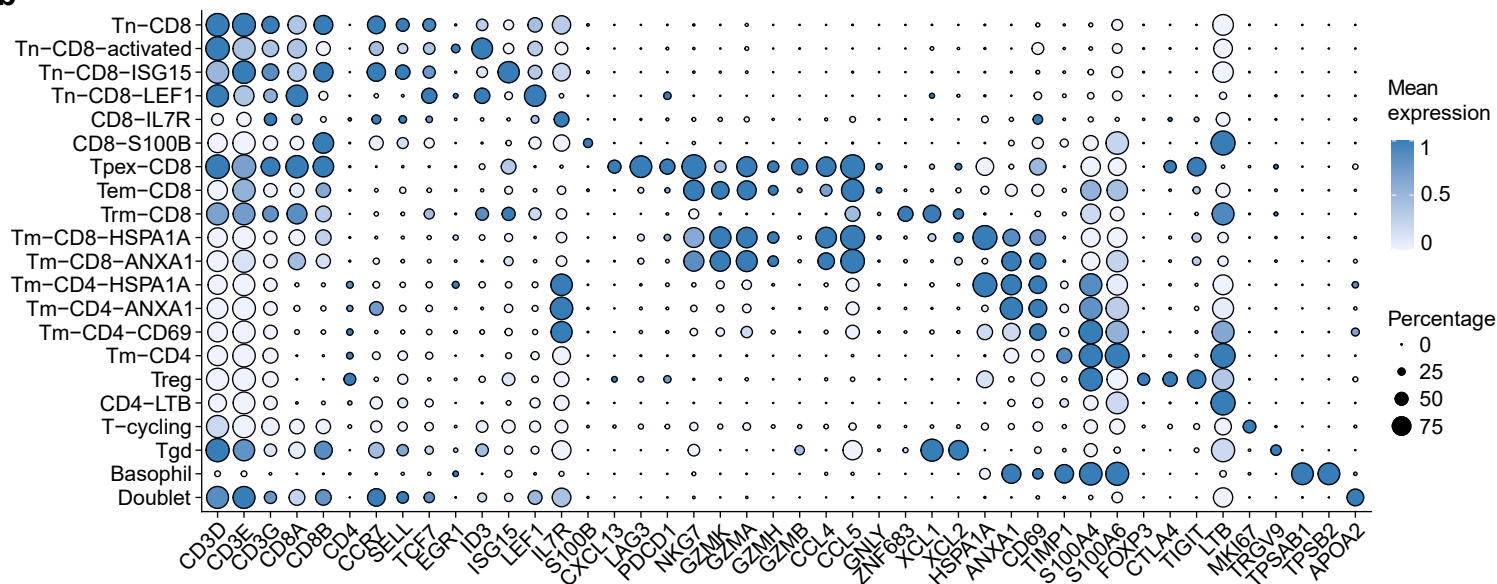
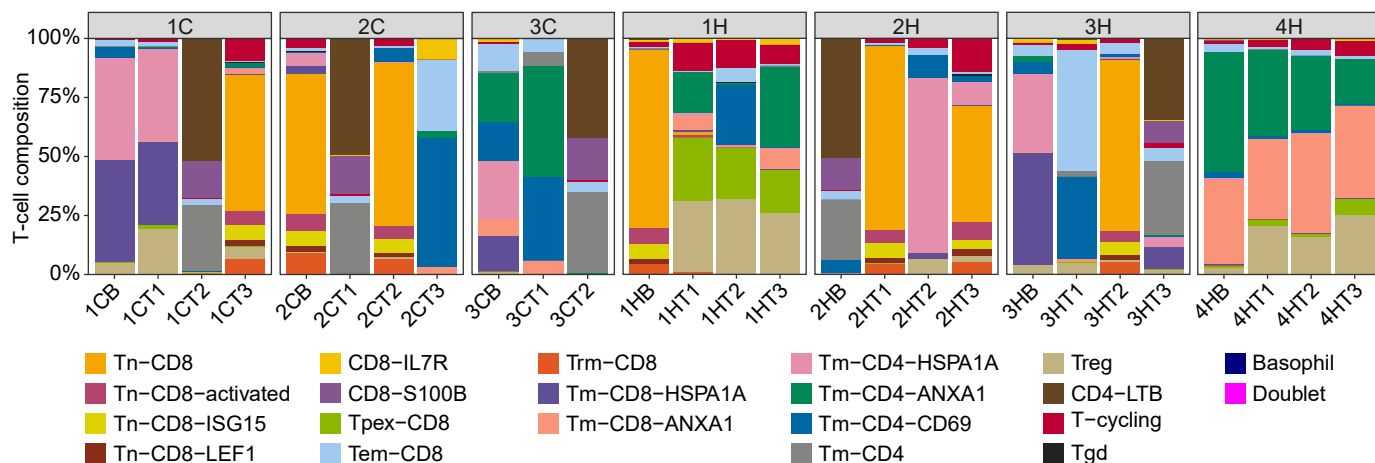
Supplementary Figure 8. Immunohistochemistry of CD3+ T cells.

(a) Representative images of CD3 staining of tumor border and tumor core regions for cases of 1C and 2H. Scale bars, 50 μ m. (b) Number of CD3+ positive cells per field of view (~ 0.2 mm). A total of 5 fields of view were randomly selected for each tumor region. Error bar, mean $\pm$ standard deviation.



Supplementary Figure 9. Analyses of non-malignant cells.

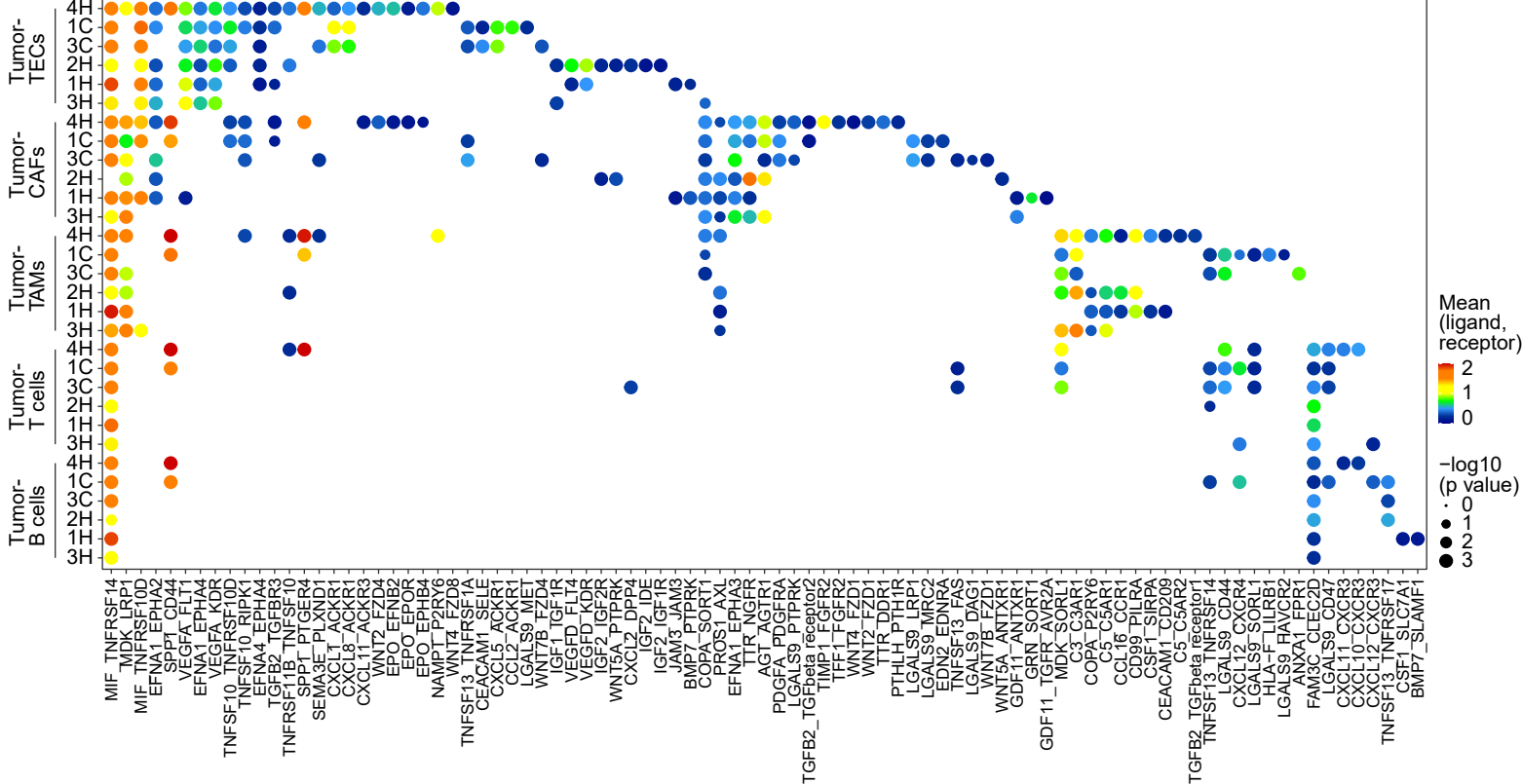
(a) Correlation of each non-malignant cell type (i.e., CAFs [n=1,626], TECs [n=1,176], TAMs [n=6,602] and B cells [n=1,774]) between different tumor regions of each individual case. Pearson's correlation test (two-sided) was applied. In the boxplots, the central rectangles span the first quartile to the third quartile, with the segments inside the rectangle corresponding to the median. Whiskers extend 1.5 times the interquartile range. (b) The relationship of tumor size and the ratio of non-malignant cell heterogeneity (average correlation value) between tumor border and tumor core. Pearson's correlation test (two-sided) was used to indicate the statistical significance of the relationship. (c) Immune score of each case. Cytotoxic, exhaustion and overall immune scores were calculated. C/E, the ratio of cytotoxic and exhaustion scores.

a**b****c**

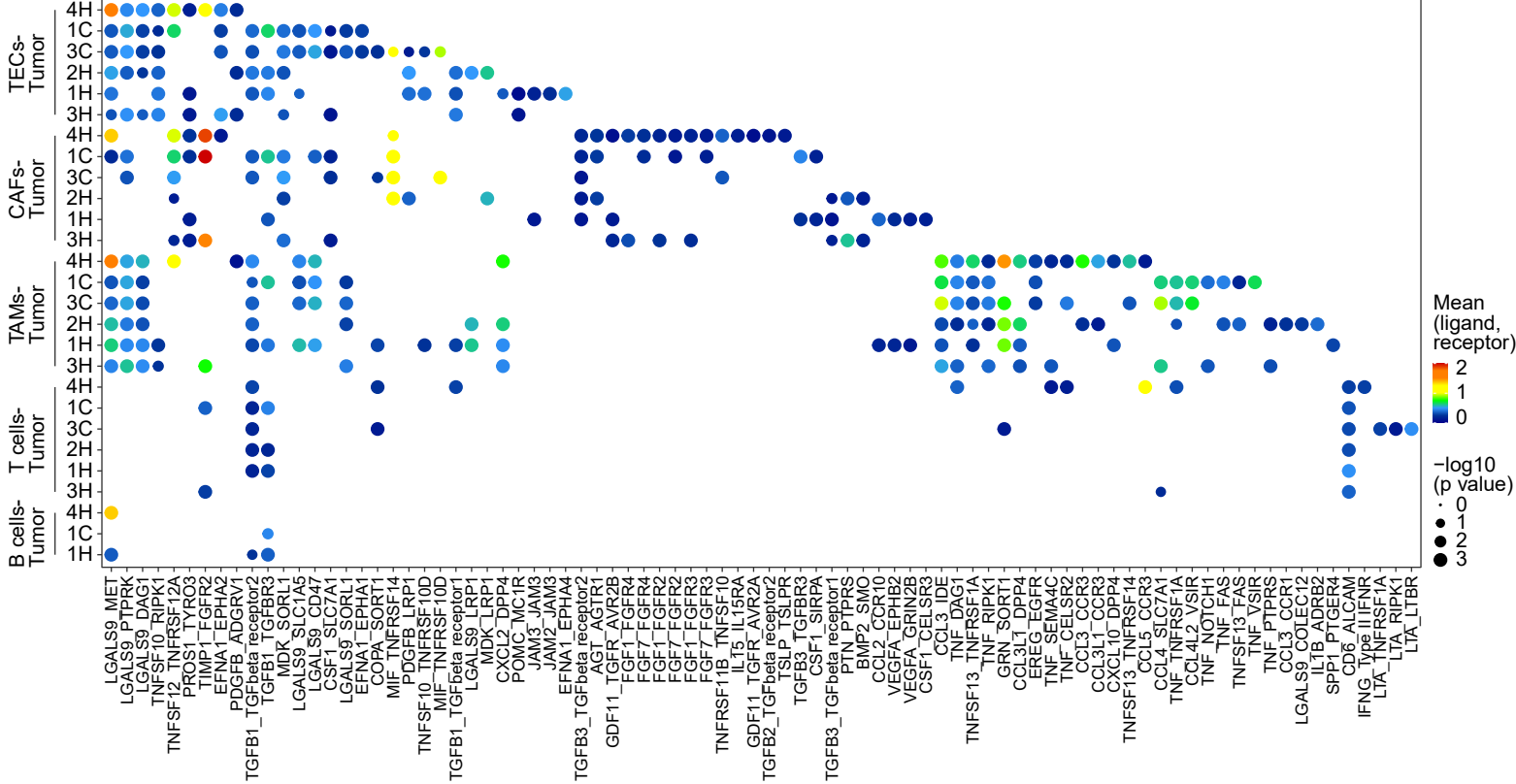
Supplementary Figure 10. T-cell subsets.

(a) UMAP of T cells colored by different subsets. Tn, native T cell; Tm, memory T cell; Tpex, pre-exhaustion T cell; Tem, effector memory T cell; Trm, tissue resident memory T cell; Treg, regulatory T cell; Tdg, gamma delta T cell. (b) The mean expression and percentage of representative genes in each T-cell subset in (a). The filled color of the circles represents the mean expression of the genes. The size of each circle indicates the proportion of cells in expressing a specific gene. (c) The composition of T-cell subsets in each tumor sample from each case. Source data are provided as a Source Data file.

Tumor-



TECs-

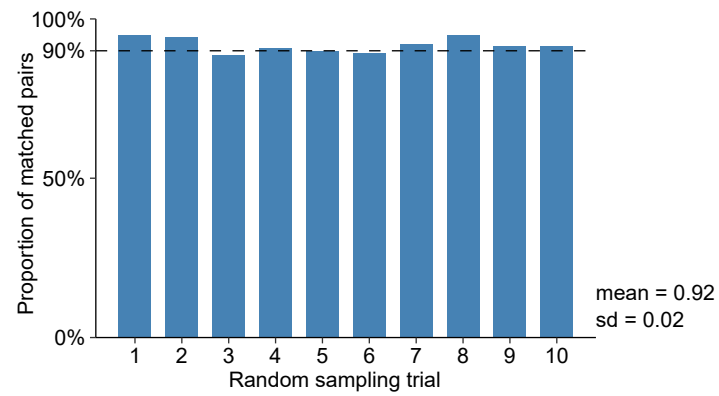


Supplementary Figure 11. Ligand-receptor interactions of malignant cells and non-malignant cells.

(a and b) Ligand-receptor interactions from direction of tumor to TME (a) and TME to tumor (b). Each column indicates a ligand-receptor pair, with the first and the second gene representing a ligand and a receptor, respectively. Each row represents an interaction between two cell types within a case. Color indicates the mean expression value of a ligand-receptor pair. The size of each dot stands for $-\log_{10}(\text{p-value})$ of a specific interaction. p-value was calculated based on 1,000 times of random permutation as described in CellPhoneDB (methods).

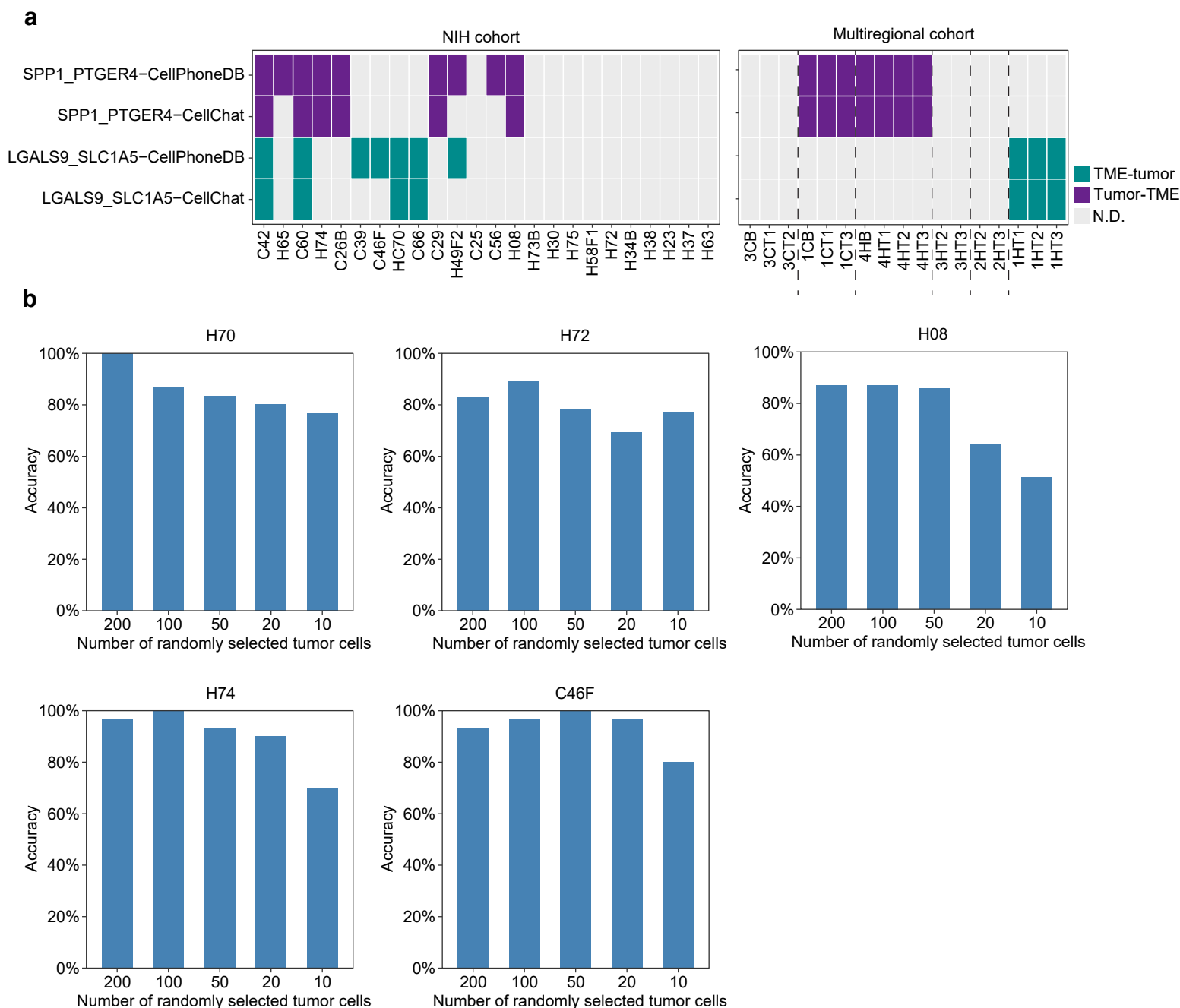
Supplementary Figure 12. Communication networks of malignant cells and non-malignant cells in multiple tumor regions of each case.

(a) Ligand-receptor interactions of malignant cells and non-malignant cells in six cases with viable malignant cells. Each row represents a tumor region. Each column indicates a ligand-receptor pair, with the first and the second gene representing a ligand and a receptor, respectively. The directions of ligand-receptor interactions are indicated by colors. Purple, malignant cells provide ligands and interact with receptors from non-malignant cells in the TME; green, non-malignant cells in the TME provide ligands and interact with receptors from malignant cells. The non-malignant cell types are indicated by colors. (b) Ligand-receptor interactions between malignant cells and T-cell subsets. Colors represent the directions of the interactions as described in (a). The size of each dot represents the proportion of tumor regions within each case in identifying a specific interaction pair, with 1 indicating occurrence in all tumor regions and 0 indicating occurrence in none of the tumor regions. (c) Stacked bar plot of the percentage of ligand-receptor pairs (between malignant cells and T-cell subsets) in each individual case found in certain proportion of tumor regions. One means that a pair was found in all tumor regions within a case while zero means that a pair was found in none of the tumor regions.



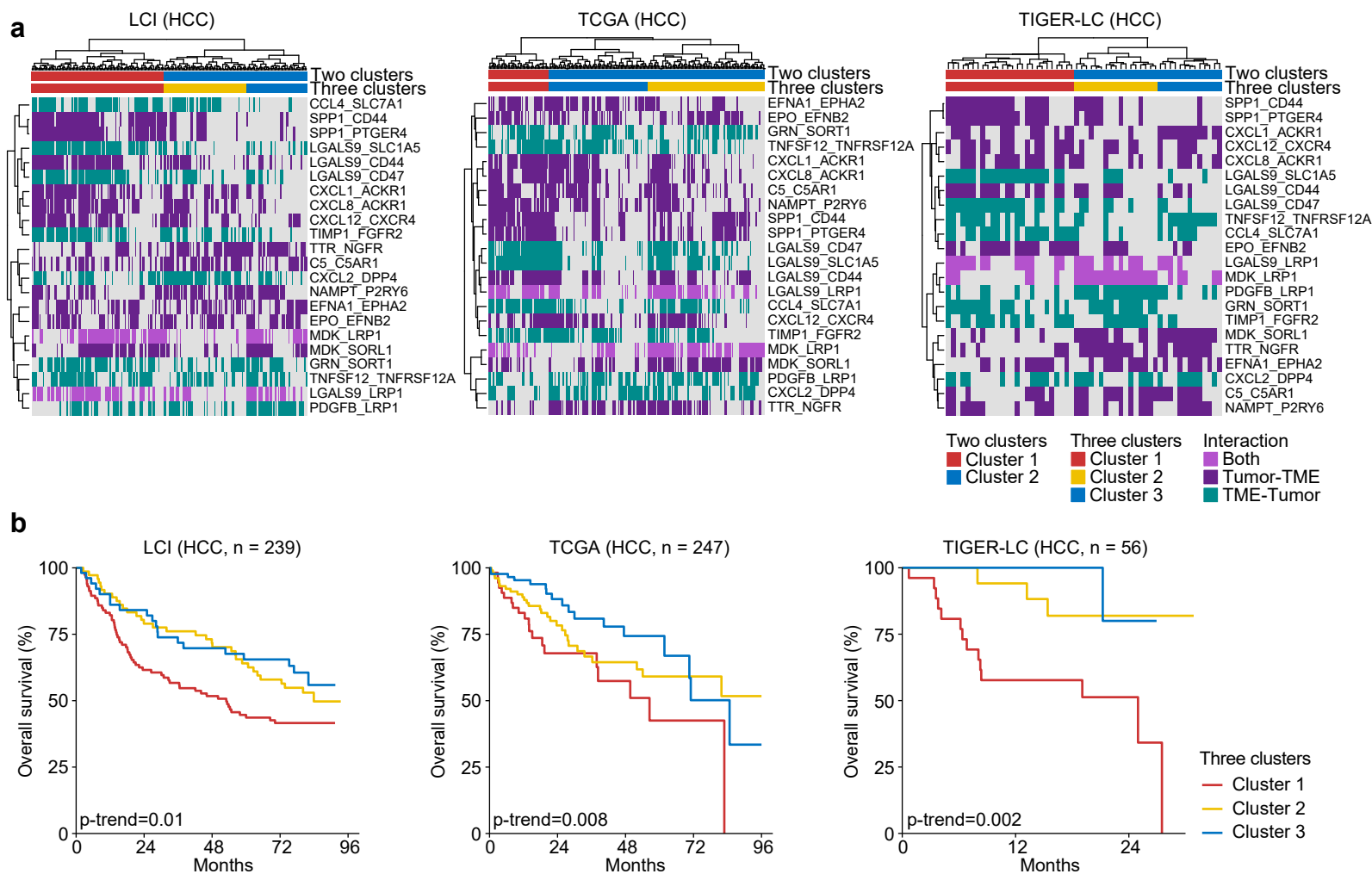
Supplementary Figure 13. Stability test by random sampling T cells and TAMs.

The proportion of matched ligand-receptor interaction pairs by using random selection of T cells and TAMs with the original search using all available single cells. sd, standard deviation.



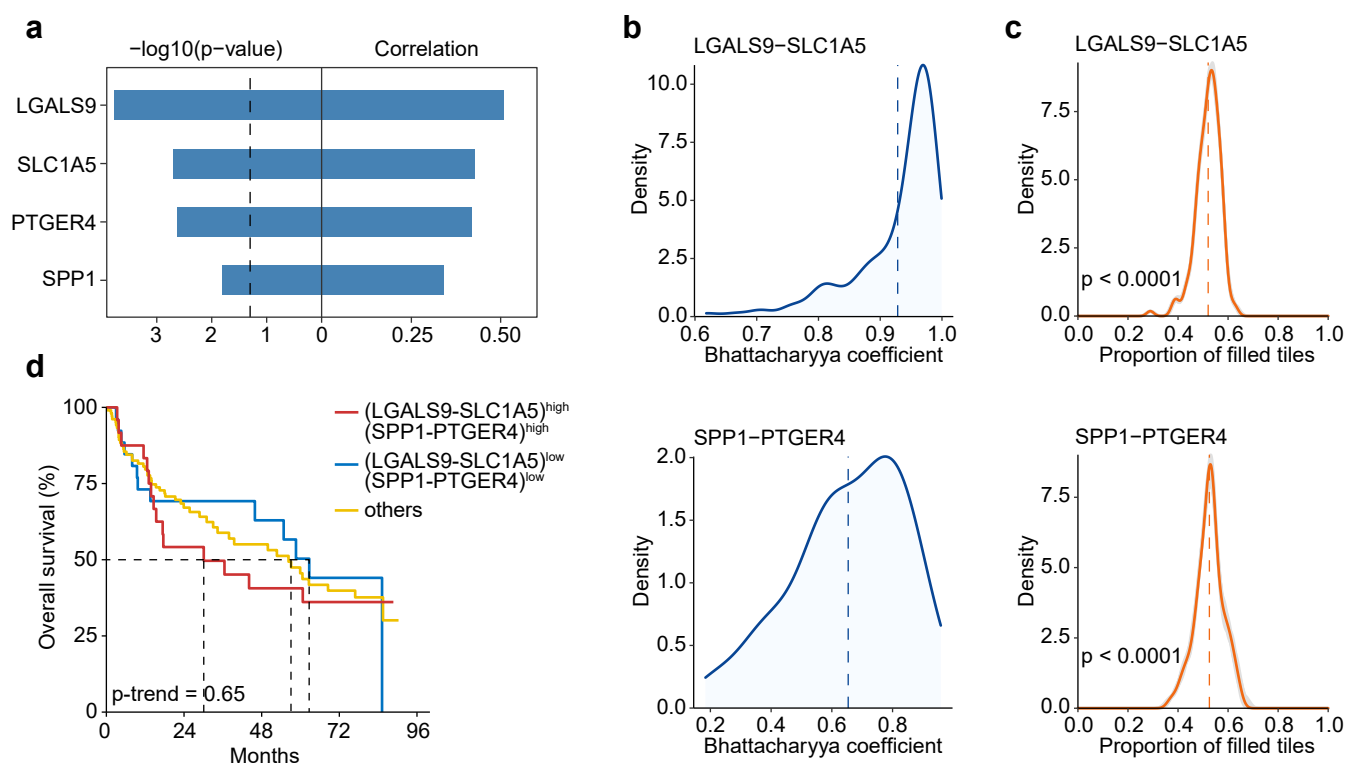
Supplementary Figure 14. Stability test of ligand-receptor pair identification.

(a) Comparison of the ligand-receptor pairs (SPP1-PTGER4, LGALS9-SLC1A5) identified using CellPhoneDB and CellChat in the NIH single-cell cohort and the multiregional single-cell cohort. (b) The accuracy of ligand-receptor pair identification by randomly sampling 200, 100, 50, 20, 10 malignant cells in each case. Accuracy was determined based on five times of random sampling for each setting and further compared with the pairs identified using the original number of malignant cells in Fig. 5a.



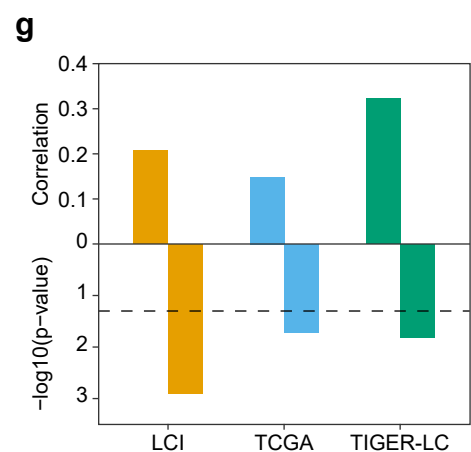
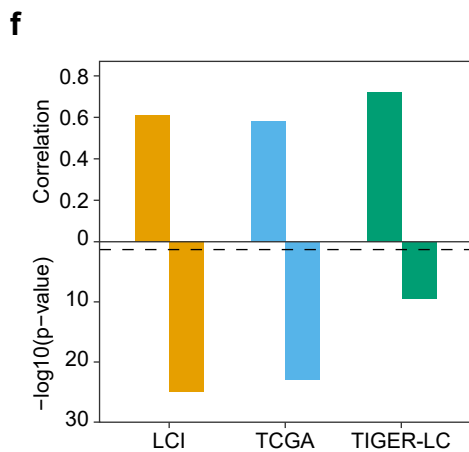
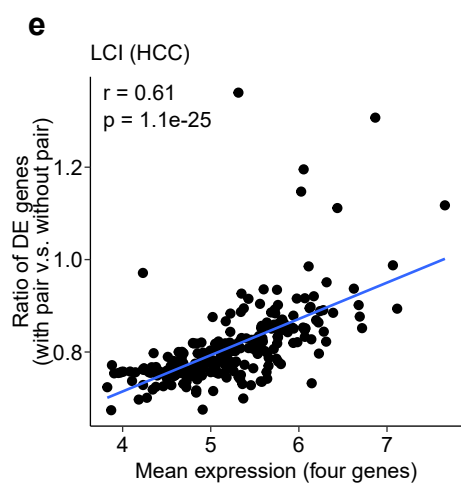
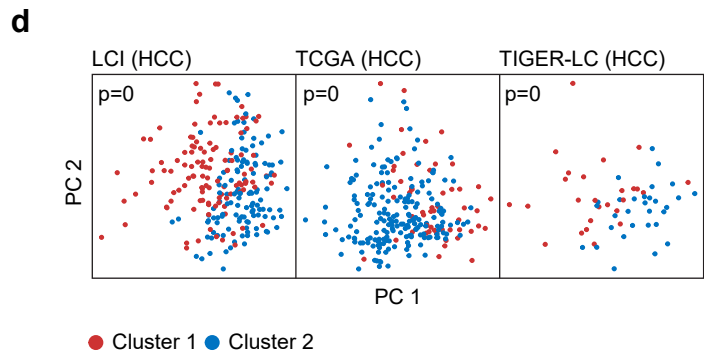
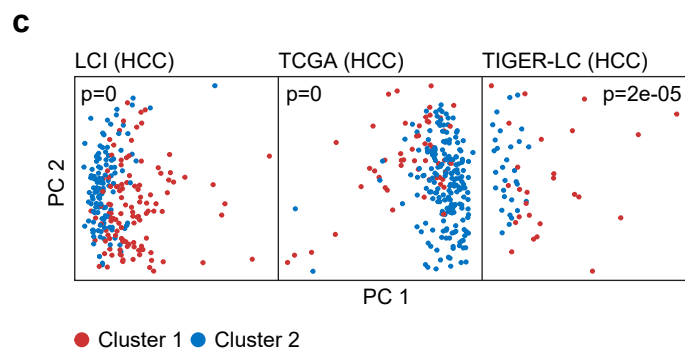
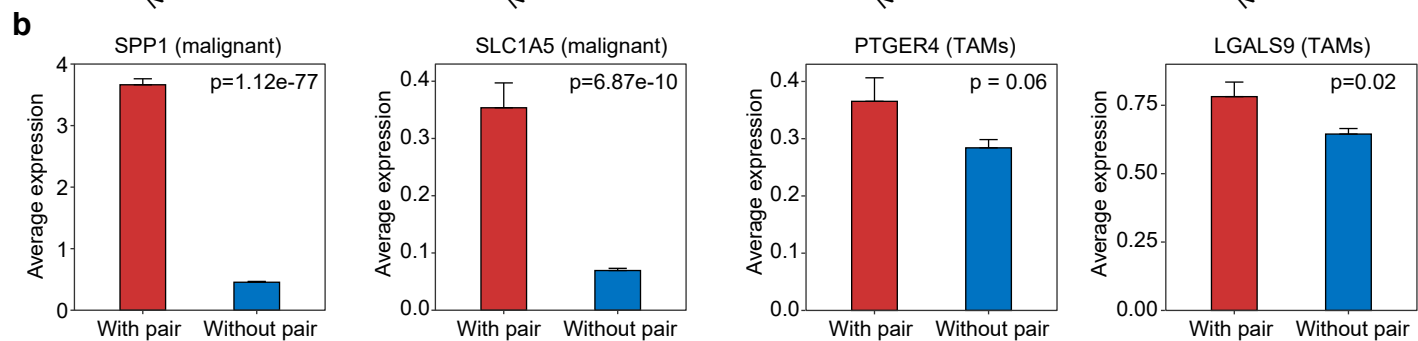
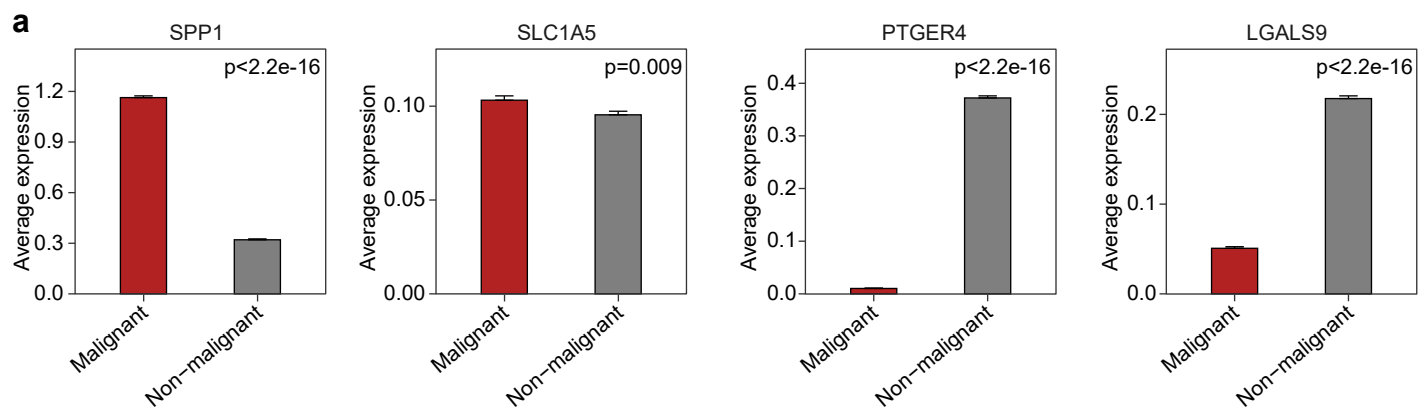
Supplementary Figure 15. Validation of the association between interaction patterns and tumor aggressiveness using bulk transcriptomic data.

(a) Hierarchical clustering of ligand-receptor interaction patterns in three HCC cohorts, i.e., LCI, TCGA, and TIGER-LC. Each row represents a ligand-receptor pair. Each column stands for a tumor sample. The directions of ligand-receptor interactions are indicated by colors. Purple, malignant cells provide ligands and interact with receptors from non-malignant cells in the TME; green, non-malignant cells in the TME provide ligands and interact with receptors from malignant cells; light purple, both directions. Tumor samples were divided into two or three clusters by cutting the hierarchical tree into two or three branches. (b) Overall survival of HCC patients from Cluster 1, Cluster 2 and Cluster 3 determined in (a). The number of samples in each cohort was provided. Log-rank test for trend was performed to show the statistical significance of the survival trend.



Supplementary Figure 16. RNAscope validation of two interaction pairs in LCI cohort.

(a) Correlation of RNAscope signal and bulk transcriptome gene expression in LCI cohort. Pearson's correlation coefficient (two-sided) was calculated. Dashed line indicates the cutoff of p value in this study, i.e., $-\log_{10}(p\text{ value}=0.05)$. (b) The distribution of Bhattacharyya coefficients of LGALS9 and SLC1A5 (top) as well as SPP1 and PTGER4 (bottom) in HCC samples from LCI cohort. Dashed line indicates the mean value. (c) The distribution of the proportion of filled tiles of LGALS9 and SLC1A5 (top) as well as SPP1 and PTGER4 (bottom) in HCC samples from LCI cohort. Randomization was used to generate the random spread of markers on tissue sections as a reference (see Methods). Gray line, proportion determined based on the ratio of true signal and each random spread; gold line, mean derived from ten gray lines. Dashed line indicates the mean value. p-value is provided to show the statistical difference from 1. (d) Overall survival of HCC patients in LCI cohort with low expression and high expression of LGALS9 and SLC1A5 as well as SPP1 and PTGER4. Tumor samples with the expression of the four marker genes in between were grouped into others. Log-rank test for trend was preformed to show the statistical significance of the survival trend.



Supplementary Figure 17. Analysis of the two pair (LGALS9-SLC1A5, SPP1-PTGER4) related signaling.

(a) Average expression of SPP1, SLC1A5, PTGER4 and LGALS9 in malignant cells (n=17,164) and non-malignant cells (n=31,154). p value was provided to indicate the statistical significance. Error bar indicates the standard error of the mean. (b) Average expression of SPP1 and SLC1A5 in malignant cells (n=4,235) as well as PTGER4 and LGALS9 in TAMs (n=1,381) in the samples with or without the two pairs (LGALS9-SLC1A5, SPP1-PTGER4) in Fig. 5a. p value was provided to indicate the statistical significance. Error bar indicates the standard error of the mean. (c and d) PCA of tumor samples in three HCC cohorts based on differentially expressed genes of malignant cells (c) or TAMs (d) derived from the samples with or without the two pairs (LGALS9-SLC1A5, SPP1-PTGER4). Cluster 1 and Cluster 2 of the three cohorts were determined in Supplementary Fig.15. Hotelling's T-squared test was used to test the difference of the multivariate means of Cluster 1 and Cluster 2. (e) Correlation analysis in LCI cohort based on the mean expression of the four genes (LGALS9, SLC1A5, SPP1 and PTGER4) and the ratio of average expression of differential expression (DE) genes derived from malignant cells in Fig. 7k. Pearson correlation coefficient (two-sided) and p value are provided. (f) Correlation coefficient and p value in all three HCC cohorts using the same analysis in (e). Dashed line indicates the cutoff of p value in this study, i.e., $-\log_{10}(p \text{ value}=0.05)$. (g) Correlation coefficient and p value in all three HCC cohorts based on similar analysis in (e) but with the DE genes in Fig. 7j. Pearson's correlation coefficient (two-sided) was calculated. Dashed line indicates the cutoff of p value in this study, i.e., $-\log_{10}(p \text{ value}=0.05)$.

Supplementary Table 1. Clinical information of seven primary liver cancer patients involved in this multiregional single-cell study.

Variable	Multiregional single-cell cohort (N=7)
Age ^a — year	
Median	72
Range	63-77
Sex — no. (%)	
Female	3 (42.9)
Male	4 (57.1)
Etiology — no. (%)	
HCV	1 (14.3)
NAFLD	1 (14.3)
NASH	3 (42.9)
None	2 (28.6)
Diagnosis — no. (%)	
HCC	4 (57.1)
iCCA	3 (42.9)
Stage — no. (%)	
I	3 (42.9)
II	3 (42.9)
IVA	1 (14.3)
Vascular invasion — no. (%)	
No	4 (57.1)
Yes	3 (42.9)
Tumor size ^b — cm	
Median	7.5
Range	3-24.5
CA19-9 — no. (%)	
≤35 U/mL	4 (57.1)
>35 U/mL	2 (28.6)
Missing data	1 (14.3)
Alpha-fetoprotein — no. (%)	
Negative, ≤20 ng/mL	4 (57.1)
Positive, >20 ng/mL	2 (28.6)
Missing data	1 (14.3)

^aAge at time of tissue collection used for this study.

^bTumor size determined by radiologic measurements.