

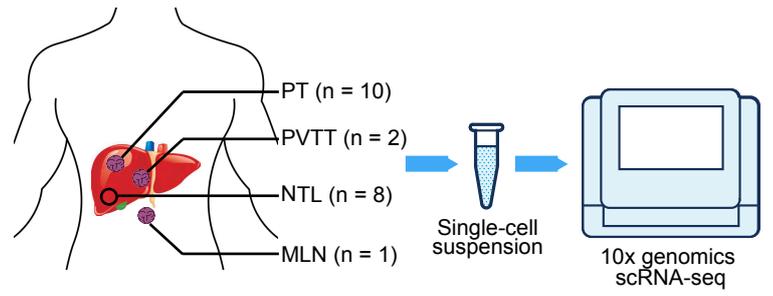
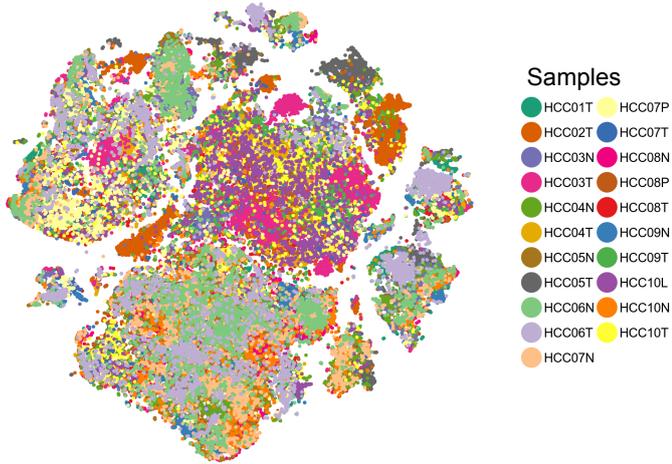
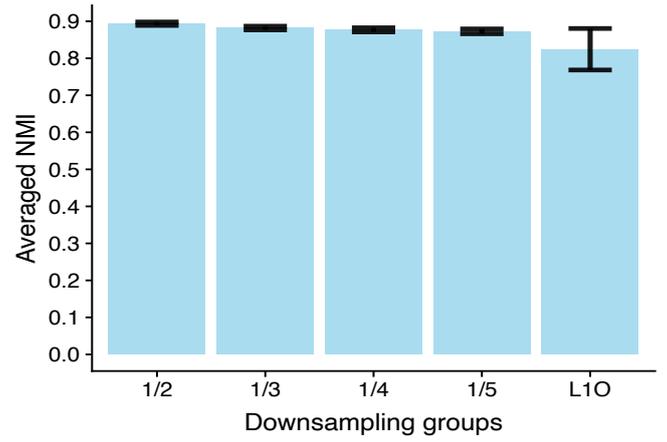
**A Single-Cell Atlas of the Multicellular Ecosystem of Primary and Metastatic  
Hepatocellular Carcinoma**

**This PDF file includes:**

Supplementary Figures 1-15 (legends included)

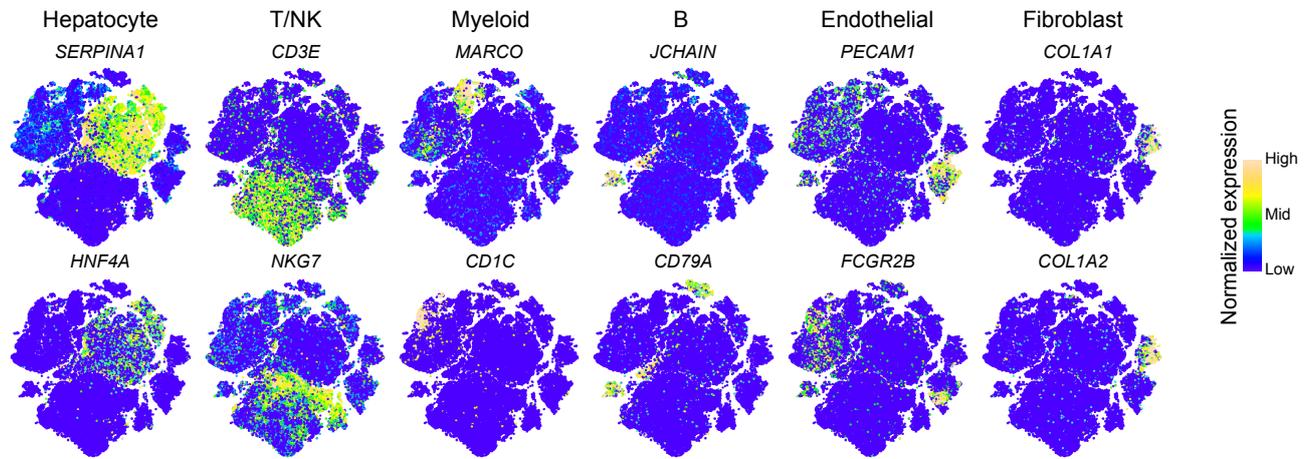
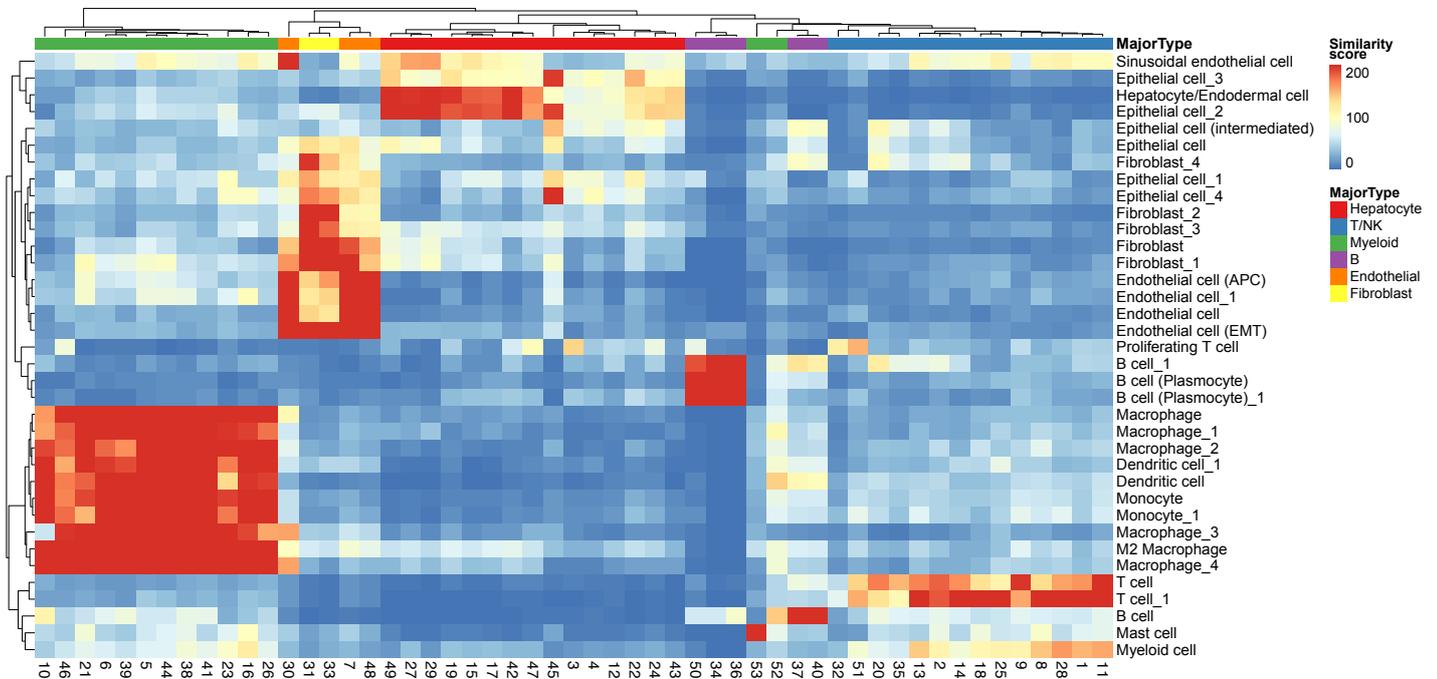
**a**

Patients	NTL	PT	PVTT	MLN	Viral infection	Stages
HCC01	No	Yes	No	No	HBV	I
HCC02	No	Yes	No	No	HBV	I
HCC03	Yes	Yes	No	No	HCV	I
HCC04	Yes	Yes	No	No	HCV	II
HCC05	Yes	Yes	No	No	None	IIIA
HCC06	Yes	Yes	No	No	HBV	IIIA
HCC07	Yes	Yes	Yes	No	None	IIIB
HCC08	Yes	Yes	Yes	No	None	IIIB
HCC09	Yes	Yes	No	No	HBV	IV
HCC10	Yes	Yes	No	Yes	HBV	IV

**b****c****d**

**Supplementary Figure 1 | Single-cell RNA sequencing of primary and metastatic HCCs, and non-tumor liver tissues.**

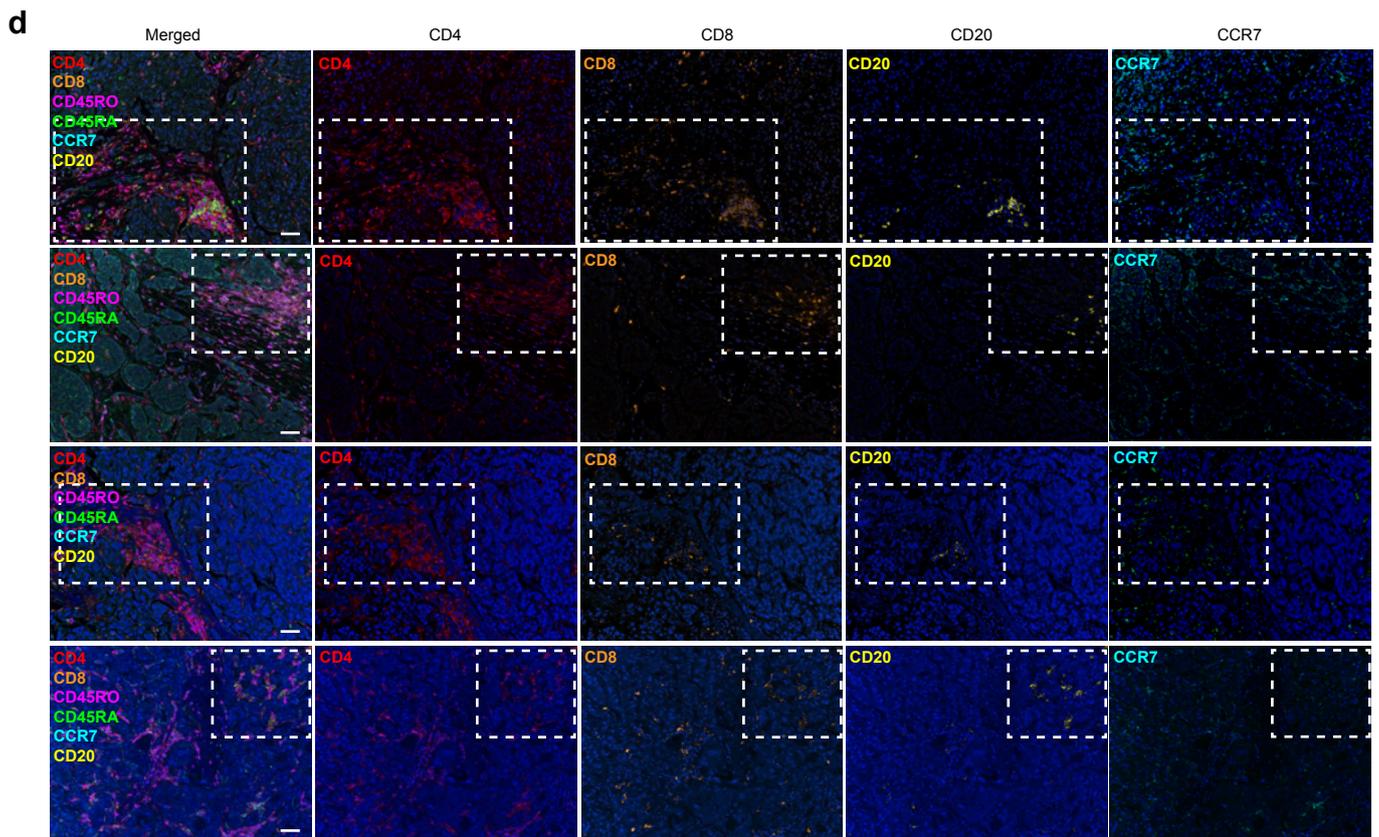
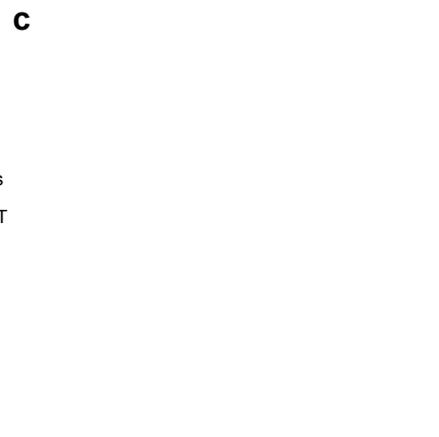
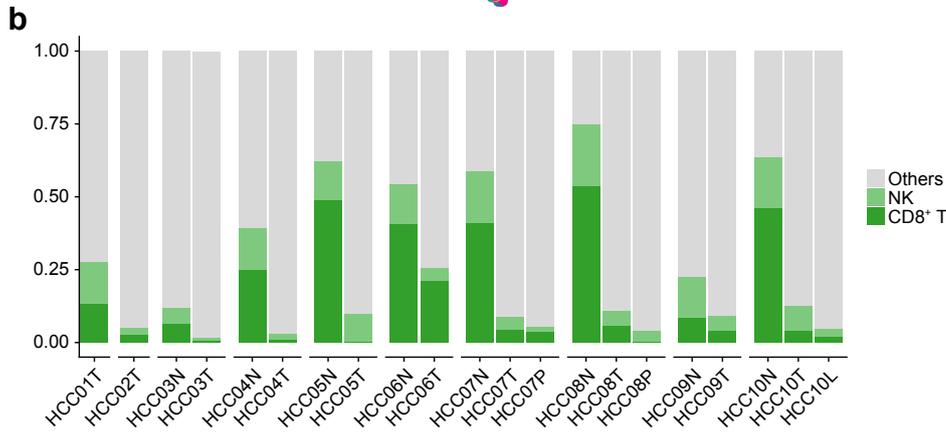
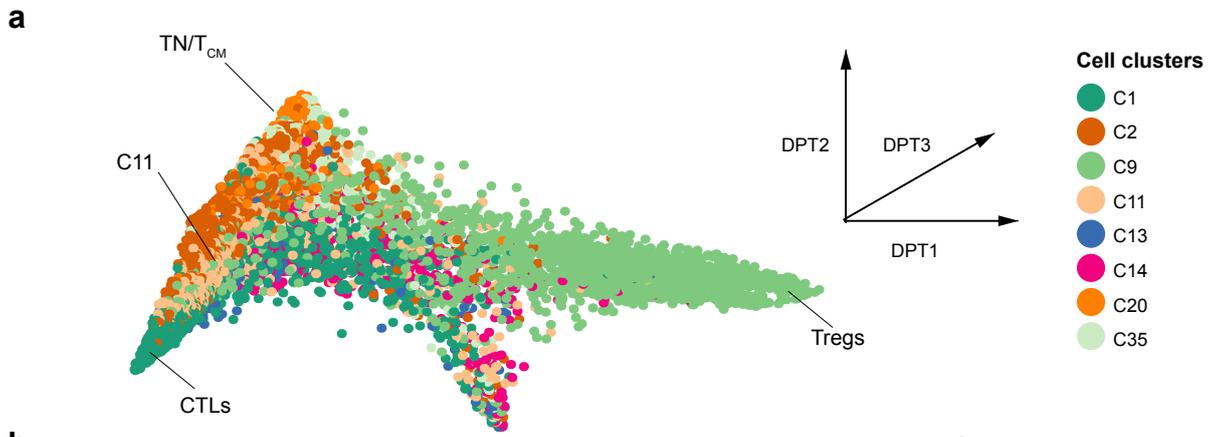
**a** Summary of the patients and their sequenced samples. More details are provided in Supplementary Data 1. NTL, non-tumor liver; PT, primary tumor; PVTT, portal vein tumor thrombus; MLN, metastatic lymph node. **b** Graphical workflow of experiment strategy. **c** The T-distributed Stochastic Neighbor Embedding (tSNE) plot of all cells with mutual nearest neighbor (MNN)-based batch correction, colored by samples. **d** The normalized mutual information (NMI) index between the clustering results of benchmark dataset and 100 times down-sampling datasets, as well as leave-one-patient-out datasets (n = 100 randomly down-sampling replicates for each group). 1/2, 1/3, 1/4 and 1/5 indicate the fractions of down-sampling. L1O, leave-one-patient-out. Error bars, mean  $\pm$  sd. Source data are provided as a Source Data file.

**a****b****c**

## **Supplementary Figure 2 | Cell type annotation of HCC cellular**

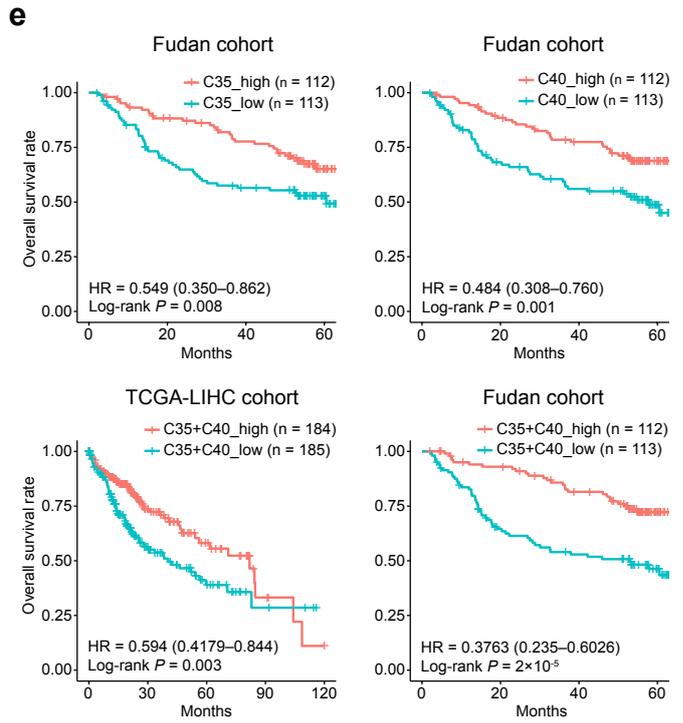
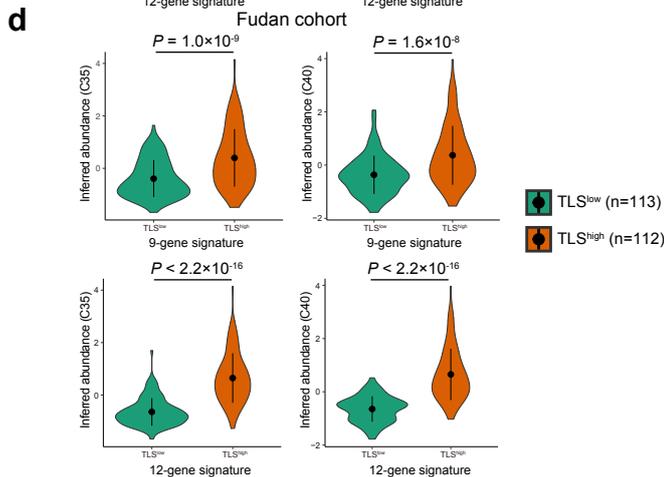
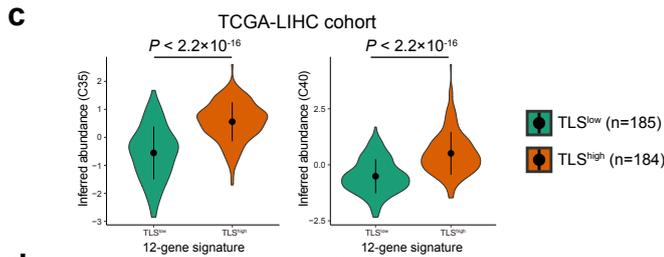
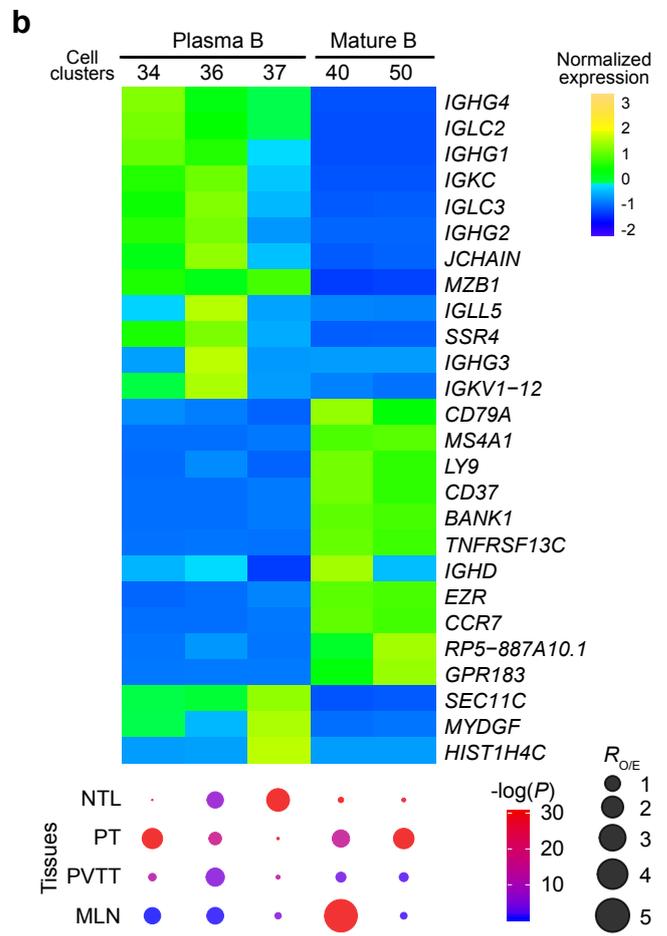
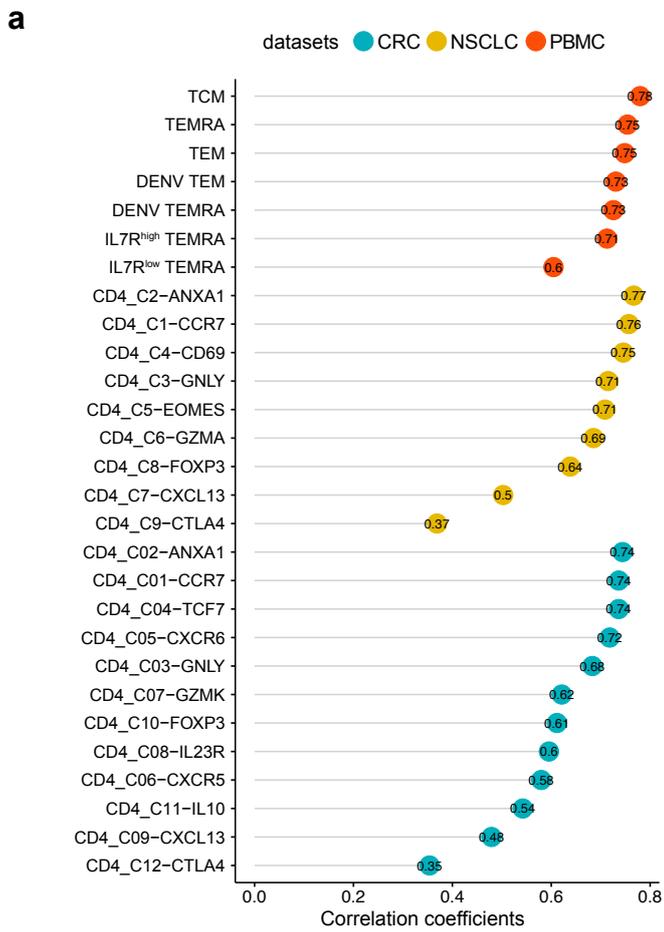
### **microenvironment.**

**a** Heatmap showing the expression profiles of upregulated genes in six major cell types. Canonical marker genes are listed at the right side of the map. Rows represent the upregulated genes. Columns represent the individual cells. **b** The tSNE plots showing the expression levels of signature genes of six major cell types, colored by gene expression. Mid, middle. **c** Heatmap showing the similarity scores between 53 cell clusters and annotated cell types in the Human Cell Landscape (HCL) project. Only liver- or HCC-related cell types were included in the analysis. Source data are provided as a Source Data file.



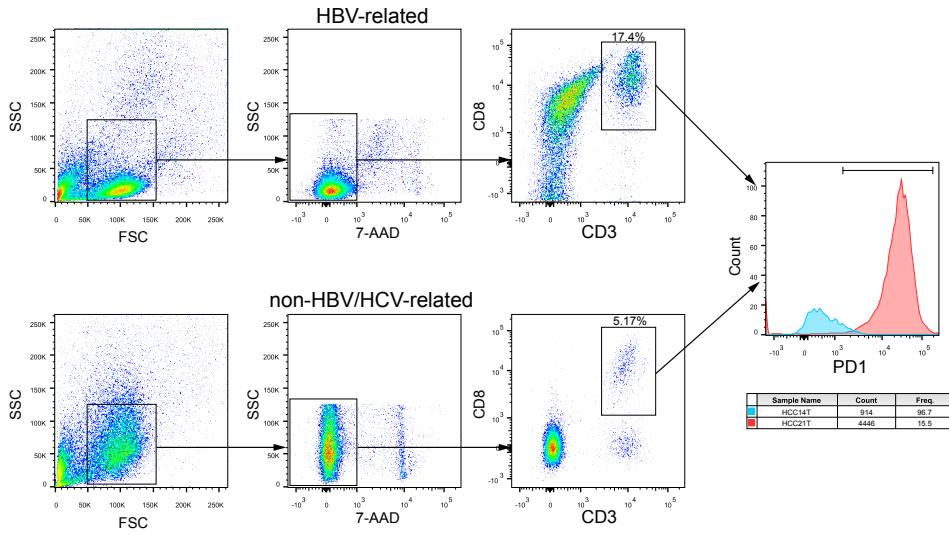
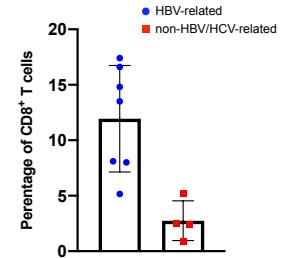
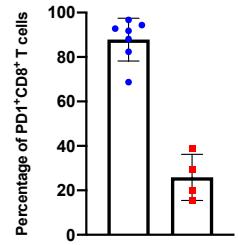
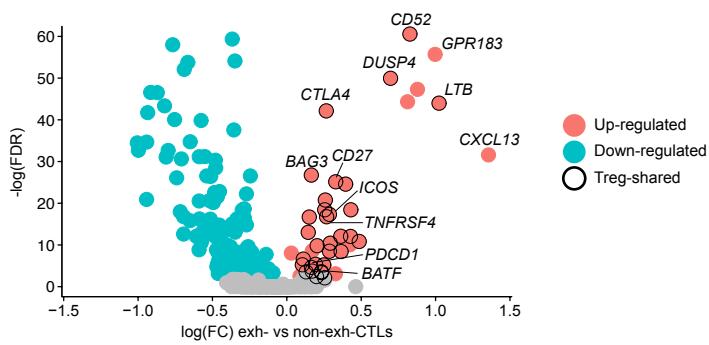
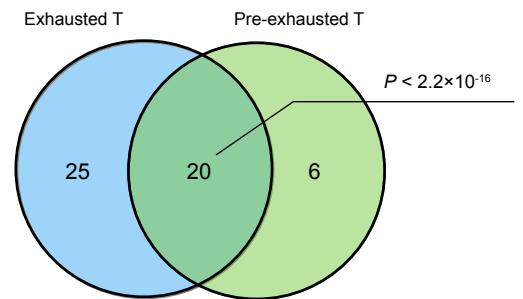
**Supplementary Figure 3 | Characterization of T cells in the cellular microenvironment of HCC.**

**a** Diffusion maps of T cells showing C11 cluster is related to an intermediate state between the TN/T<sub>CM</sub> cells (C20 and C35) and cytotoxic T lymphocytes (CTLs) (C1). Three diffusion components (DCs) are indicated by arrows. Cells of different T cell clusters are marked by colors. **b** The fractions of CD8<sup>+</sup> T and natural killer (NK) cells in primary/metastatic tumor and non-tumor liver tissues. **c** Density of CD4<sup>+</sup>CCR7<sup>+</sup>CD45RA<sup>+</sup> naïve T (TN) and CD4<sup>+</sup>CCR7<sup>+</sup> CD45RO<sup>+</sup> central memory T (T<sub>CM</sub>) cells in paired primary tumor and non-tumor liver tissues (n = 9 HCC independent patients). Among the nine HCC patients, six patients (HCC03, HCC04, HCC05, HCC06, HCC08 and HCC09) have been performed scRNA-seq on their tissue samples, and three (HCC11, HCC12 and HCC13) were the additional recruited HCC patients. Cell density was obtained using InForm (PerkinElmer) software. Boxplot elements are defined in the Methods section (section on data visualization). **d** Multicolor immunohistochemistry (IHC) staining to validate the enrichment of CD4<sup>+</sup> T<sub>CM</sub> (center) in primary tumors, exemplified by patients HCC03, HCC04, HCC05, HCC06 and HCC08, respectively. The scale bars represent 50 μm. Source data are provided as a Source Data file.



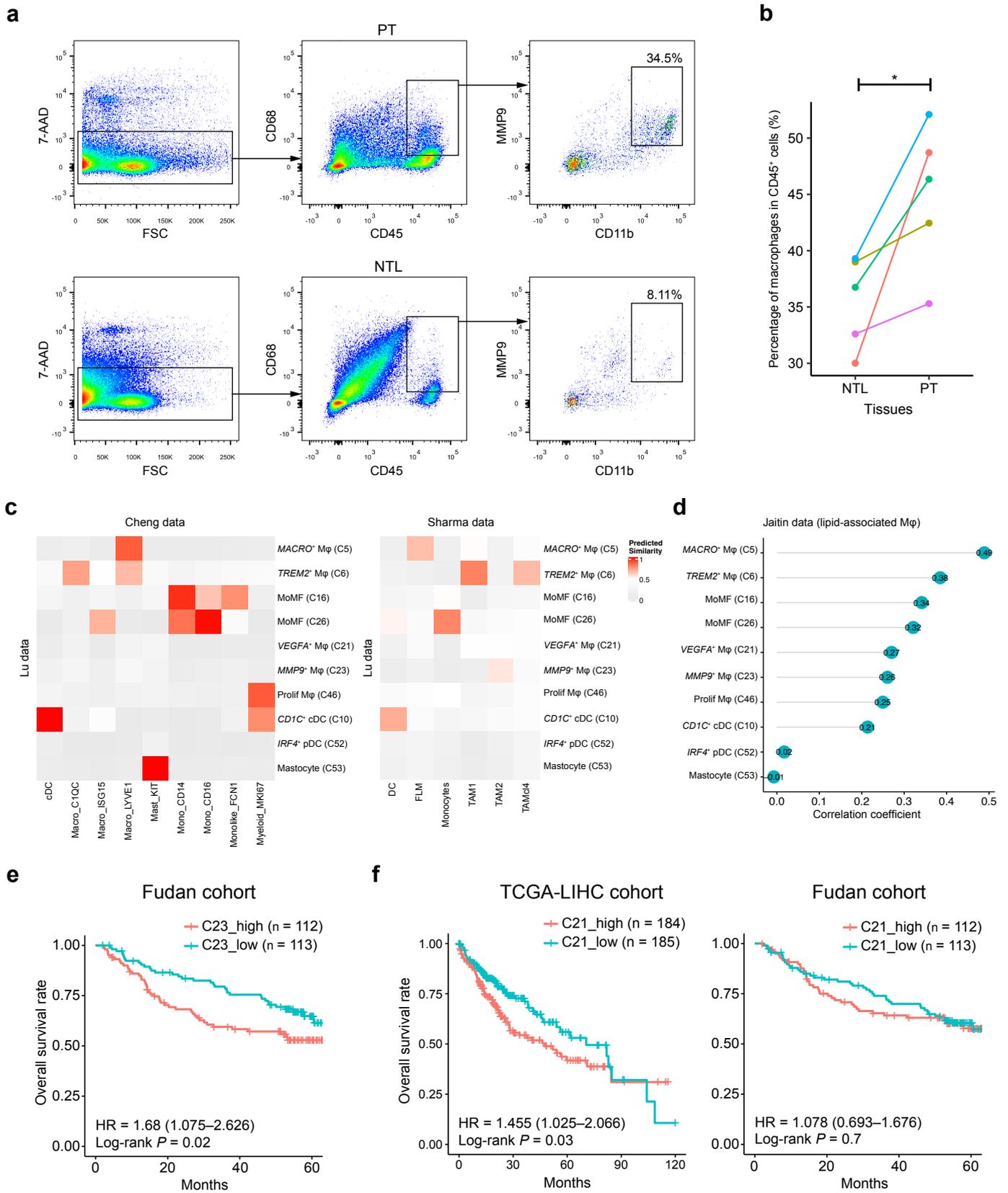
**Supplementary Figure 4 | Identification of intratumoral central memory T cells by multicolor immunohistochemistry assays.**

**a** Cluster similarity between CD4<sup>+</sup> central memory T (T<sub>CM</sub>) cells in C35 and T cell subtypes identified in peripheral blood mononuclear cells (PBMC), non-small-cell lung cancer (NSCLC) and colorectal cancer (CRC). **b** Heatmap (top) showing the expression profiles of canonical marker genes of B cell subtypes and dotplot (bottom) showing their tissue preferences. Dot size indicates the ratios of the observed *versus* expected cell numbers ( $R_{O/E}$ ); Dot color indicates the log-transformed  $P$  values determined by two-sided Chi-squared test. **c** and **d** Differential abundances of CD4<sup>+</sup> central memory T (T<sub>CM</sub>; C35) and CD20<sup>+</sup> B cells (C40) between the TLS<sup>low</sup> and TLS<sup>high</sup> HCC tumors. The bulk RNA-seq datasets of HCC tumors from The Cancer Genome Atlas (TCGA)-liver hepatocellular carcinoma (LIHC) cohort (b) (n = 369) and the Fudan HCC cohort (c) (n = 225) were used in the analyses. The HCC tumors were classified as TLS<sup>low</sup> or TLS<sup>high</sup> tumors based on a 9-gene TLS signature and a 12-gene TLS signature using median split, respectively. The abundances of cell subtypes were inferred from the bulk RNA-seq data using the signature genes of respective subtype. The statistical significance was determined by unpaired two-tailed  $t$  test. TLS, tertiary lymphoid structures. Source data are provided as a Source Data file.

**a****b****c****d****e**

**Supplementary Figure 5 | T cell exhaustion in HBV-/HCV-related and non-HBV/HCV-related HCC tumors.**

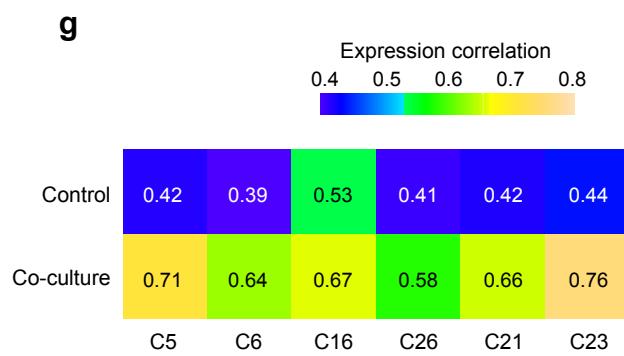
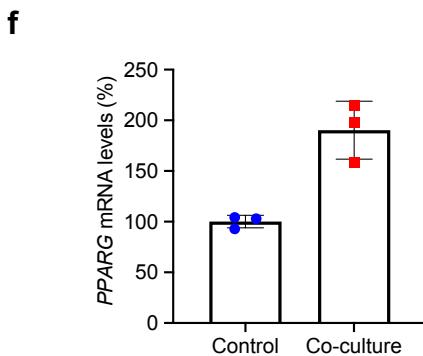
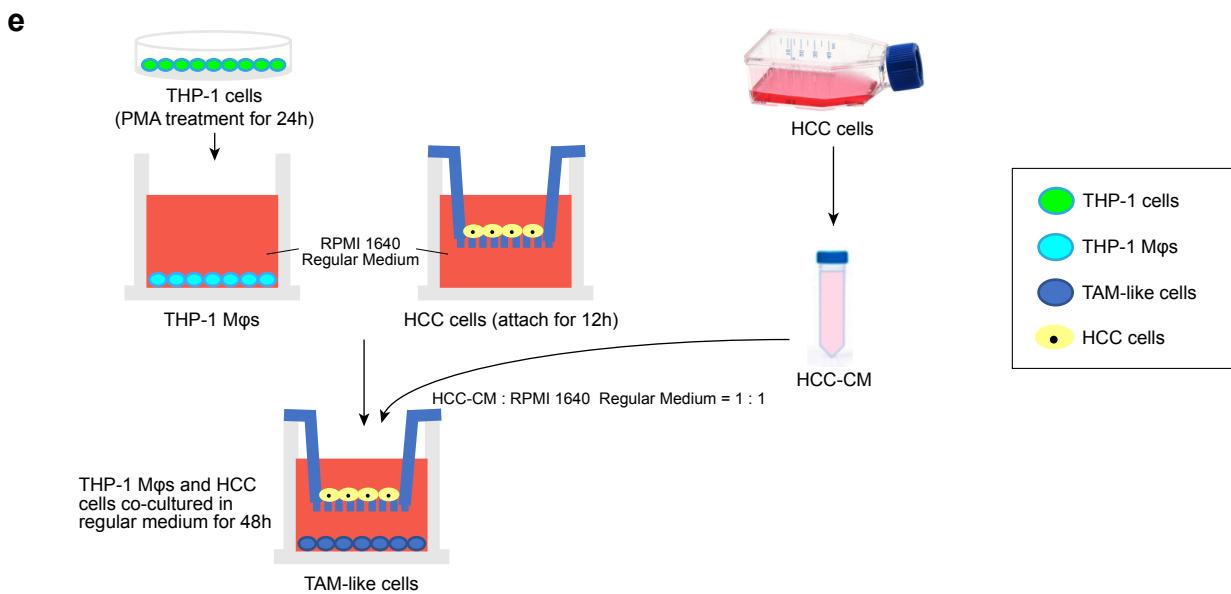
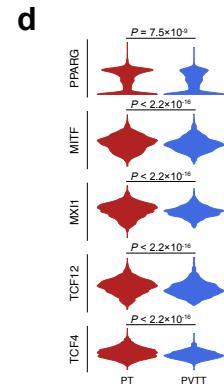
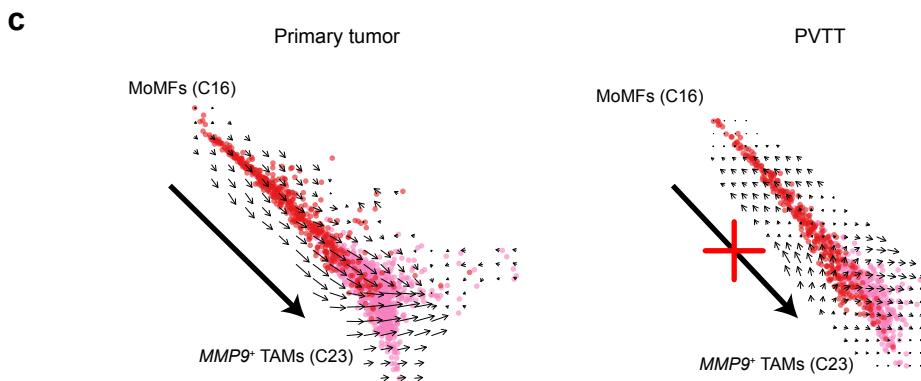
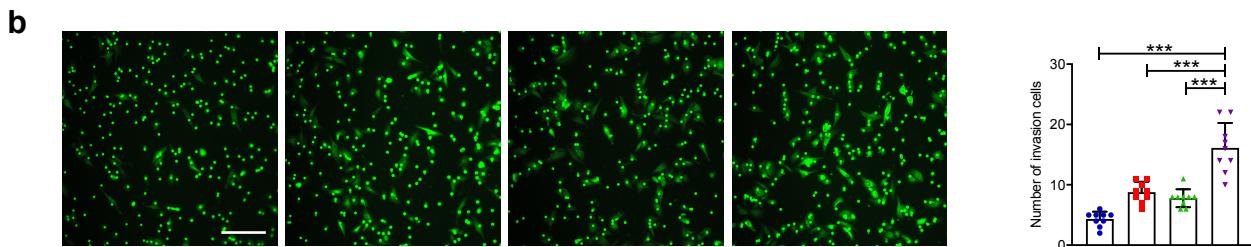
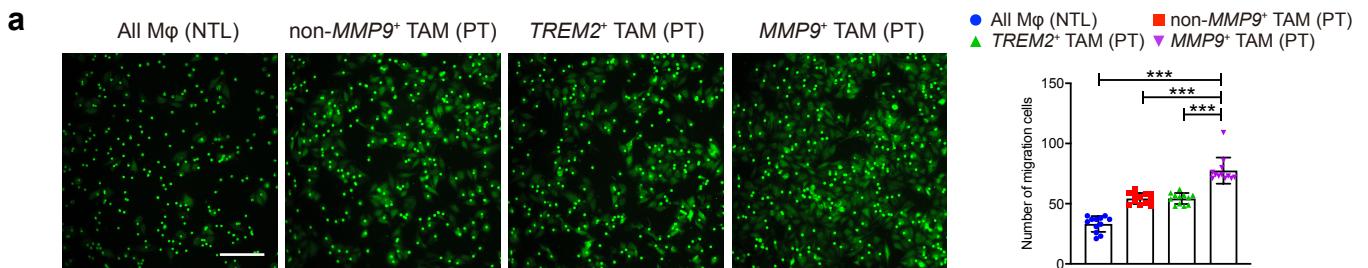
**a** Fluorescence-activated cell sorting (FACS) gating strategy of PD1<sup>+</sup> exhausted CD8<sup>+</sup> T cells sorted from primary tumor (PT) tissues in HBV-/HCV-related and non-HBV/HCV-related HCC tumors. Percentages of CD8<sup>+</sup> T cells in CD45<sup>+</sup> immune cells **b** and PD1<sup>+</sup> exhausted CD8<sup>+</sup> T cells in CD8<sup>+</sup> T cells **c** in primary tumor (PT) tissues in HBV-/HCV-related (n = 7 independent HCC patients) and non-HBV/HCV-related HCC tumors (n = 4 independent HCC patients). Error bars, mean ± sd. **d** Up- and down-regulated genes in the exhausted CTLs compared to non-exhausted CTLs. The significantly up-regulated genes are colored by red, the significantly down-regulated genes are colored by blue, and genes without significant expression changes are colored by grey. The genes also up-regulated in Tregs are marked by solid circles. CTLs, cytotoxic T lymphocytes. **e** Venn plot showing that the up-regulated genes (n = 45) in the exhausted CTLs (KM5) and the up-regulated genes (n = 26) in the pre-exhausted CTLs (KM4) are significantly overlapped. The statistical significance was determined by hypergeometric test. Source data are provided as a Source Data file.



**Supplementary Figure 6 | Characterization of *MMP9*<sup>+</sup> tumor-associated macrophages (TAMs) in the HCC cellular microenvironment.**

**a** Fluorescence-activated cell sorting (FACS) gating strategy of *MMP9*<sup>+</sup> TAMs from the primary tumor (PT) and non-tumor liver (NTL) tissues of HCC patient. **b** The proportion of CD45<sup>+</sup>CD68<sup>+</sup> macrophages in CD45<sup>+</sup> immune cells in primary tumors (PT) is significantly higher than in non-tumor liver (NTL) tissues, measured by FACS in five HCC patients. The statistical significance was determined by two-sided paired Student's *t* test. **c** Heatmaps showing cluster similarities between patient-shared myeloid clusters identified in this study and myeloid cell subtypes identified in Cheng's pan-cancer (left) and Sharma's HCC (right) datasets. **d** Lollipop plot showing similarities between lipid-associated macrophages in Jaitin's study and myeloid clusters identified in this study. **e** Higher abundances of *MMP9*<sup>+</sup> TAMs (C23) in tumors predict worse overall survival (OS) rates in HCC patients from the Fudan cohort. Classification of C23\_high/C23\_low and C21\_high/C21\_low was performed by median split respectively. Hazard ratio (HR) (with 95% confidence interval in brackets) was calculated using a Cox proportional hazards regression model, and the statistical significance was determined by log-rank test. **f** The associations between the abundances of *VEGFA*<sup>+</sup> TAMs (C21) in tumors and the overall survival (OS) rates in HCC patients from both the TCGA-LIHC (left) and the Fudan cohort (right).

Source data are provided as a Source Data file.



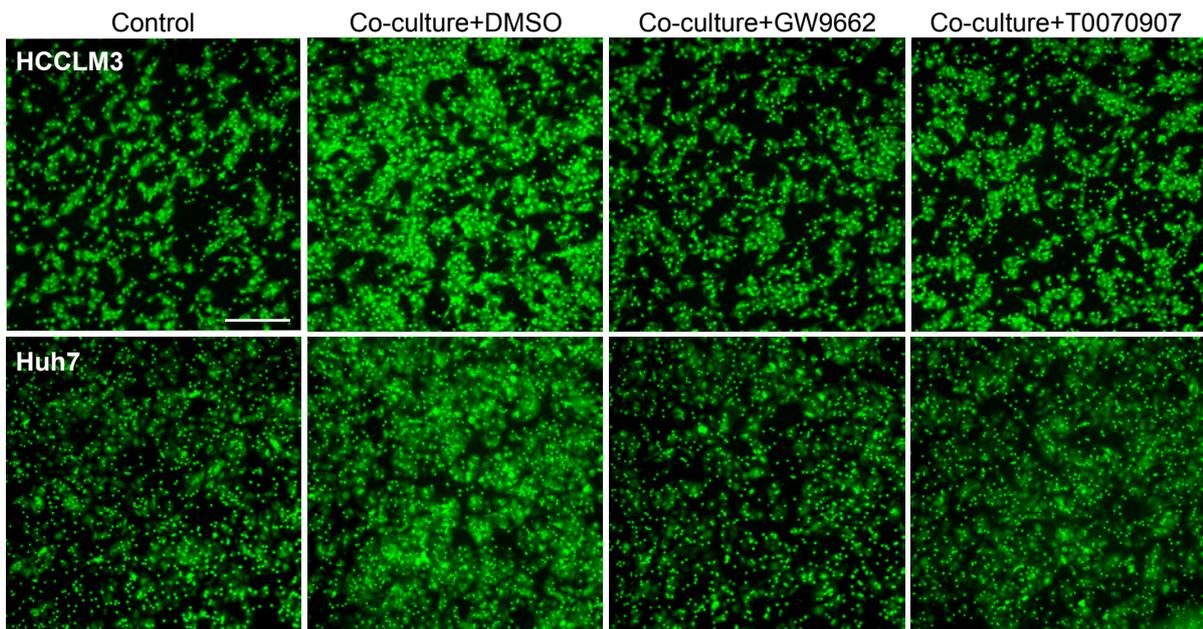
**Supplementary Figure 7 | *MMP9*<sup>+</sup> TAMs promote HCC cells invasion and migration.**

**a** and **b** HCCLM3 cells treated with *MMP9*<sup>+</sup> TAMs isolated from the PT tissues show increased abilities of **d** migration (n = 12 biological replicates) and **e** invasion (n = 9 biological replicates) compared to the control groups treated with the *TREM2*<sup>+</sup> TAMs, non-*MMP9*<sup>+</sup> TAMs isolated from the PT tissues and the whole macrophage populations isolated from the NTL tissues, respectively. The scale bars represent 20  $\mu$ m. Error bars, mean  $\pm$  sd. The statistical significance was determined by two-sided Student's *t* test. \**P* < 0.05, \*\**P* < 0.01 and \*\*\**P* < 0.001. **c** RNA velocities between the *MMP9*<sup>+</sup> (C23) and monocyte-derived macrophages (MoMFs) in C16 cluster are visualized on DPT projection in primary tumor (left) and portal vein tumor thrombus (PVTT) (right). **d** The differential activities of *MMP9*<sup>+</sup> TAM-related TFs in the *MMP9*<sup>+</sup> TAMs (C23) and MoMFs (C16) between PT and PVTT tissues. Unpaired two-tailed *t* test. **e** The coculture procedures of THP-1 cells with HCC cells. Briefly, THP-1 cells were first induced to differentiate into THP-1 macrophages by PMA treatment before the coculture. THP-1 macrophages were then cocultured with HCC cells to differentiate into TAM-like cells. **f** Expression levels of *PPARG* measured by RT-qPCR in THP-1 macrophages cultured alone (Control), cocultured with HCC cells (Co-culture) (n = 3 biological replicates). Error bars, mean  $\pm$  sd. **g** The transcriptome-wide gene expression correlations between the *MMP9*<sup>+</sup> TAMs sorted from PT by FACS and the six patient-shared macrophage clusters (C5, C6, C16, C26, C21 and C23) identified by scRNA-seq in this study. Source data are provided as a Source

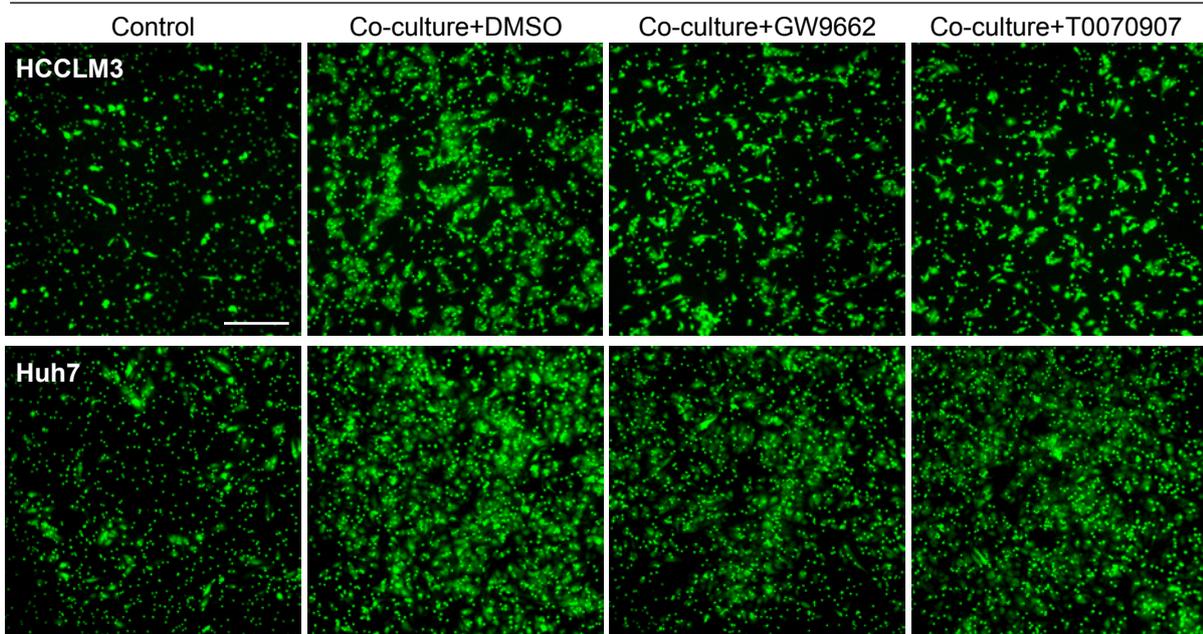
Data file.

**a**

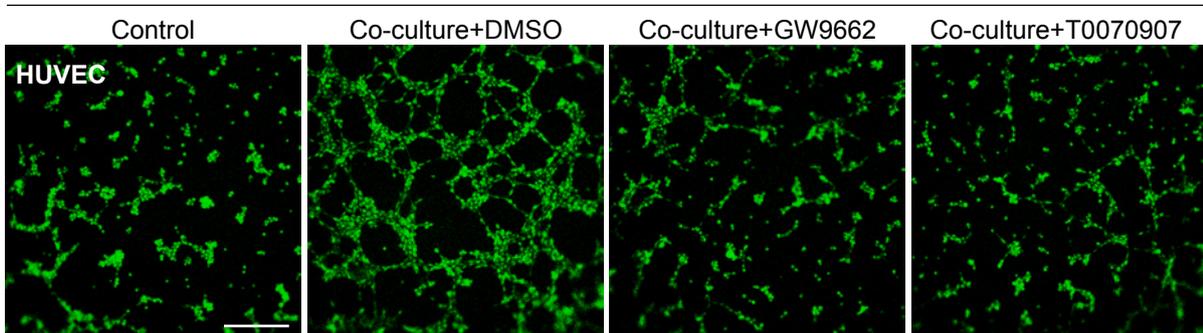
## Cell migration

**b**

## Cell invasion

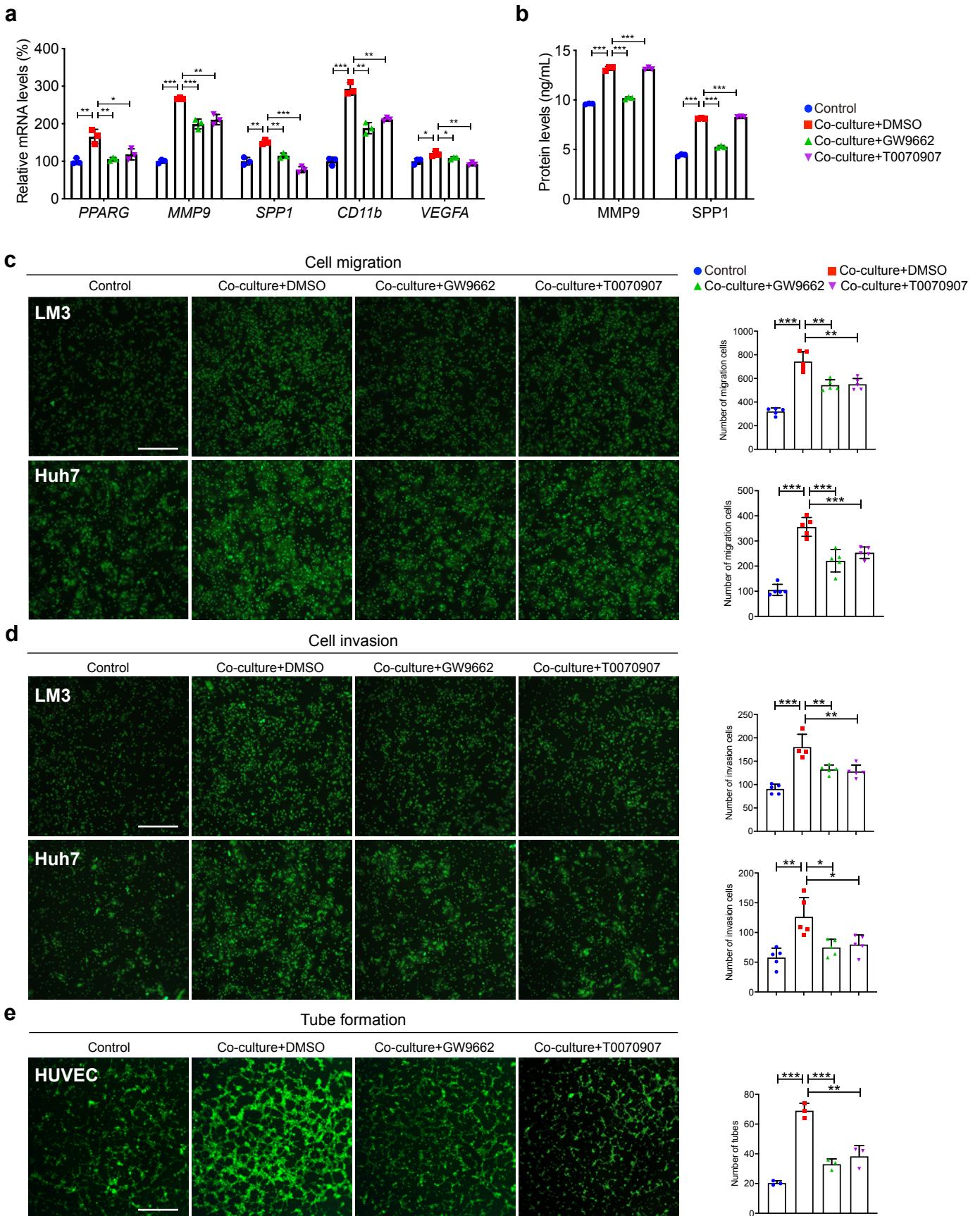
**c**

## Tube formation



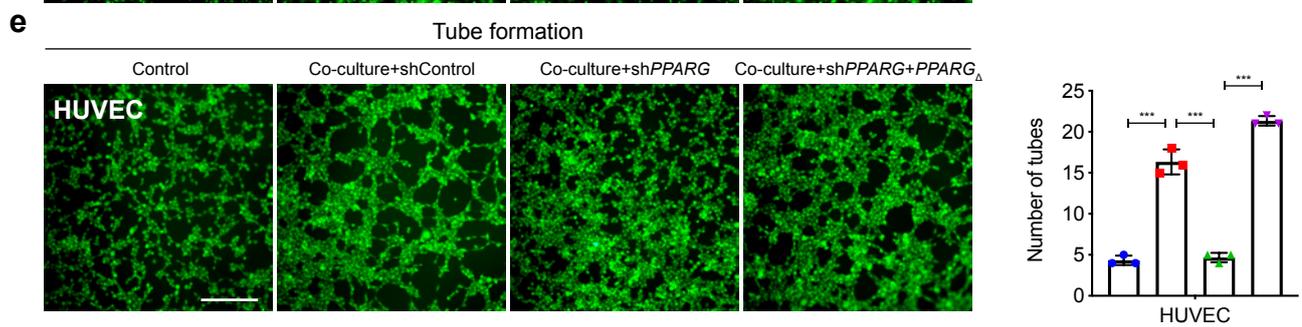
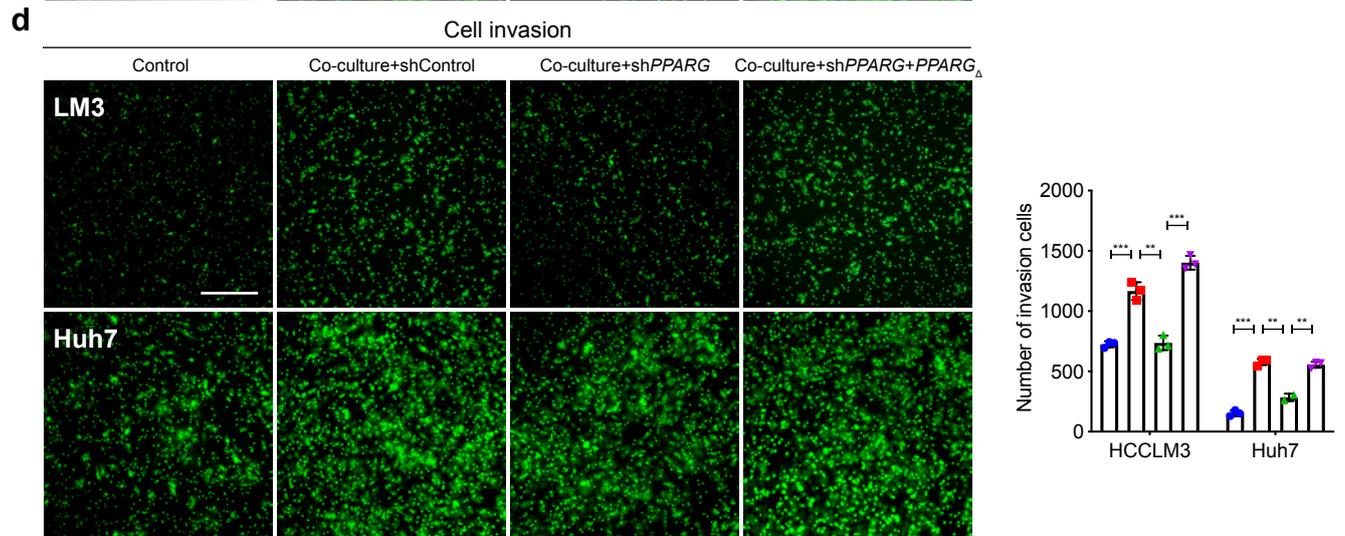
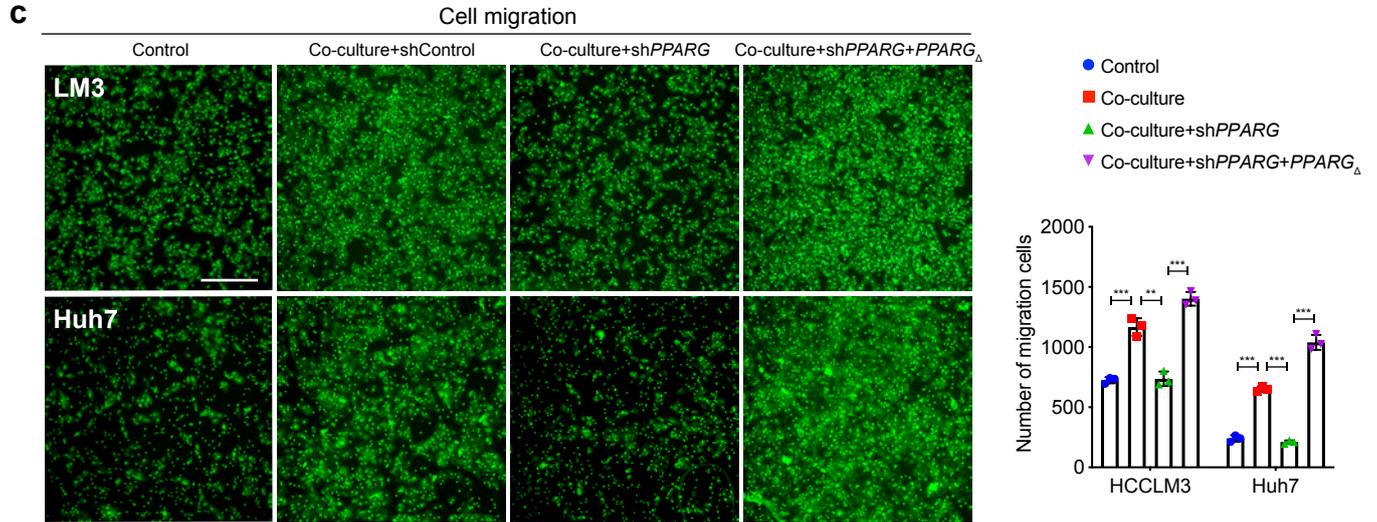
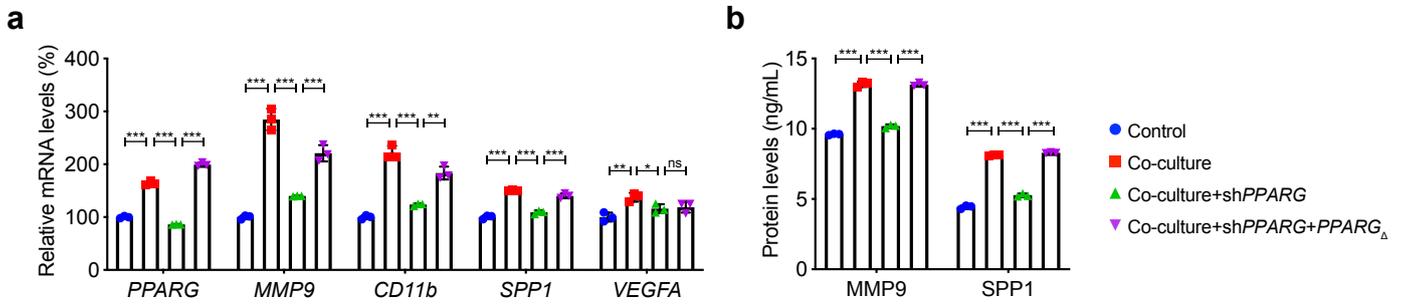
**Supplementary Figure 8 | PPAR $\gamma$  effects HCC progression through regulating *MMP9*<sup>+</sup> TAMs differentiation assessed by PPAR $\gamma$  inhibition in THP-1 cell culture.**

**a** Migration and **b** invasion abilities of HCCLM3 and Huh7 cells treated with THP-1 macrophages cultured alone (Control), cocultured with HCC cells (Co-culture+DMSO), or cocultured in the presence of the PPAR $\gamma$  inhibitors GW9662 (Co-culture+GW9662) and T0070907 (Co-culture+T0070907). **c** Tubes formation of human umbilical vein endothelial cells (HUVECs) treated with different sets of THP-1 macrophages: Control, Co-culture+DMSO, Co-culture+GW9662 and Co-culture+T0070907. In **a–c**, the experiment was repeated independently at least 3 times with similar results. The scale bars represent 20  $\mu$ m. Source data are provided as a Source Data file.



**Supplementary Figure 9 | PPAR $\gamma$  effects HCC progression through regulating *MMP9*<sup>+</sup> TAMs differentiation assessed by PPAR $\gamma$  inhibition in primary *MMP9*<sup>+</sup> TAMs culture.**

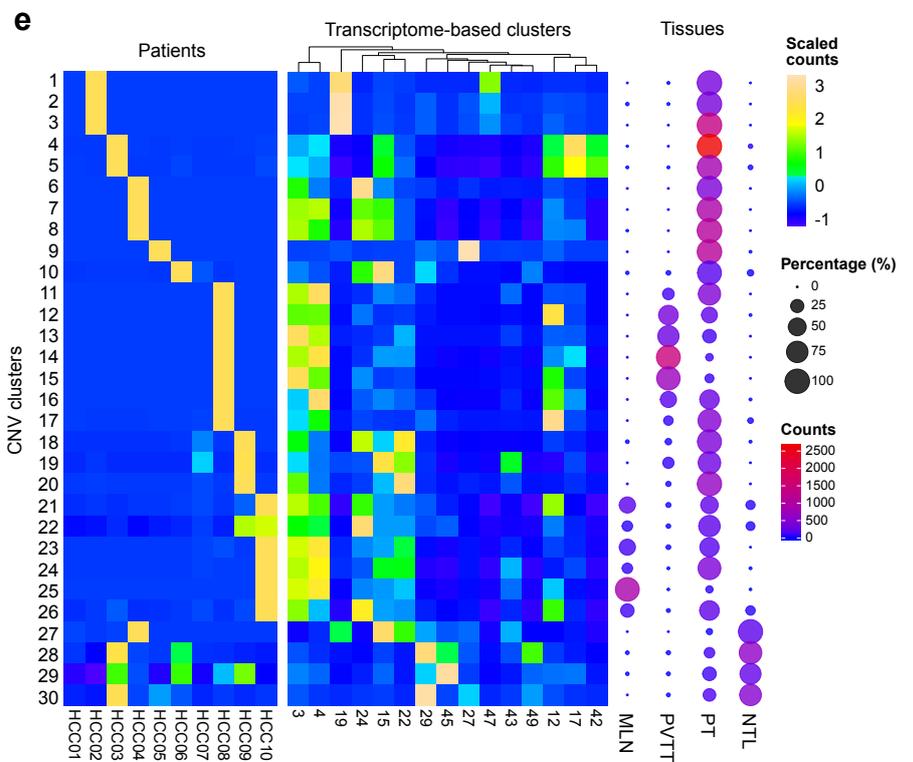
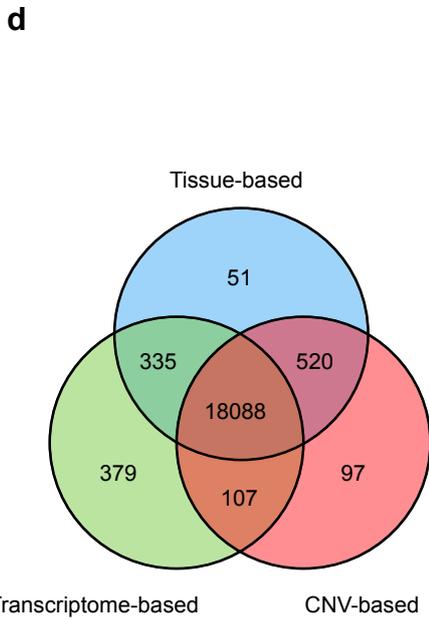
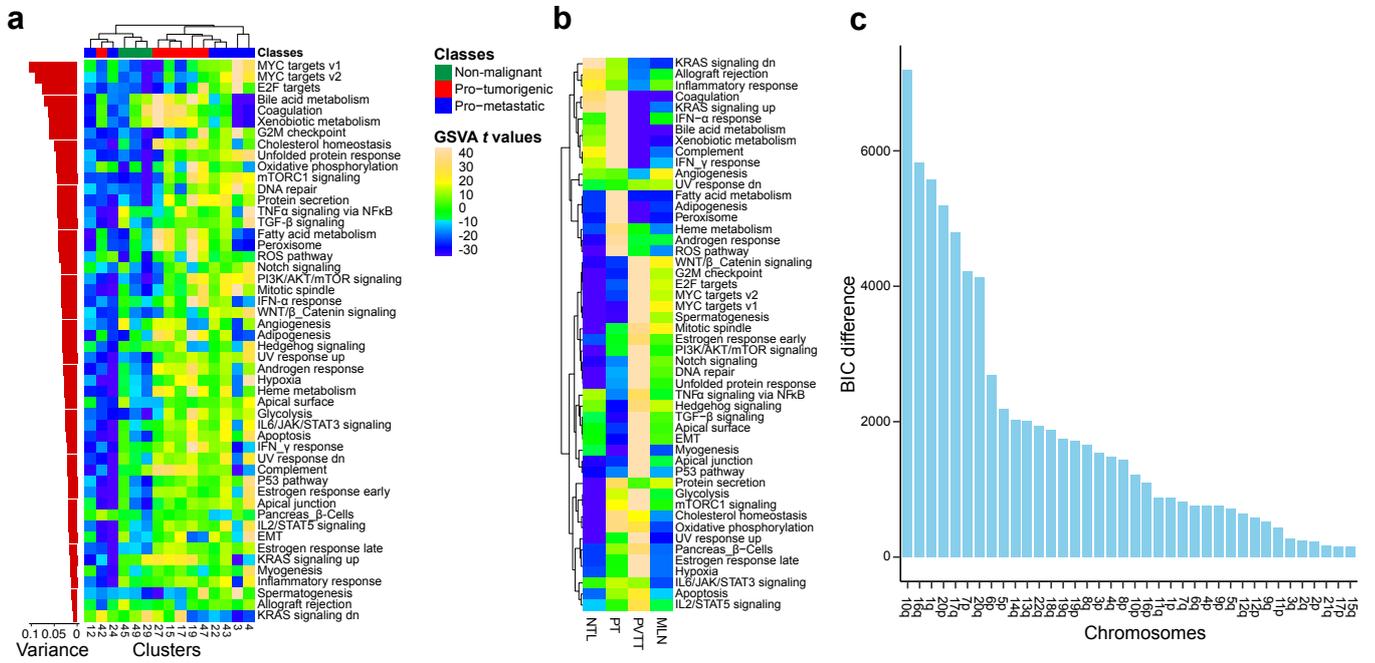
**a** Expression levels of *MMP9*<sup>+</sup> TAMs markers measured by quantitative reverse transcription real-time polymerase chain reaction (qRT-PCR) in primary *MMP9*<sup>+</sup> TAMs cultured alone (Control), co-cultured with HCCLM3 cells (Co-culture+DMSO), or co-cultured in the presence of the PPAR $\gamma$  inhibitors GW9662 (Co-culture+GW9662) or T0070907 (Co-culture+T0070907) (n = 3 biological replicates). **b** Protein levels of *MMP9*<sup>+</sup> TAMs markers MMP9 and SPP1 measured by enzyme-linked immunosorbent assay (ELISA) in the culture media of primary *MMP9*<sup>+</sup> TAMs of different groups (n = 3 biological replicates). **c** Migration and **d** invasion abilities of HCCLM3 and Huh7 cells treated with THP-1 macrophages cultured alone (Control), cocultured with HCC cells (Co-culture+DMSO), or cocultured in the presence of the PPAR $\gamma$  inhibitors GW9662 (Co-culture+GW9662) and T0070907 (Co-culture+T0070907) (n = 5 biological replicates). **e** Tubes formation of human umbilical vein endothelial cells (HUVECs) treated with different sets of THP-1 macrophages: Control, Co-culture+DMSO, Co-culture+GW9662 and Co-culture+T0070907 (n = 3 biological replicates). The scale bars represent 20  $\mu$ m. In **a–e**, error bars indicate mean  $\pm$  sd and the statistical significances were determined by Student's *t* test. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001 and \*\*\*\**P* < 0.0001. Source data are provided as a Source Data file.



**Supplementary Figure 10 | PPAR $\gamma$  effects HCC progression through regulating *MMP9*<sup>+</sup> TAMs differentiation assessed by *PPARG* knockdown in THP-1 cell culture.**

**a** Expression levels of *MMP9*<sup>+</sup> TAMs markers measured by quantitative reverse transcription real-time polymerase chain reaction (qRT-PCR) in THP-1 macrophages cultured alone (Control), cocultured with HCC cells (Coculture), or *PPARG* knockdown before the coculture (Coculture+sh*PPARG*) or rescue of *PPARG* expression in *PPARG* knockdown THP-1 macrophages before the coculture (Coculture+sh*PPARG*+sh*PPARG*<sub>SM</sub>) (n = 3 biological replicates). **b** Protein levels of *MMP9*<sup>+</sup> TAMs markers MMP9 and SPP1 measured by enzyme-linked immunosorbent assay (ELISA) in the culture media of primary *MMP9*<sup>+</sup> TAMs of different groups (n = 3 biological replicates). **c** Migration and **d** invasion abilities of LM3 and Huh7 cells treated with THP-1 macrophages cultured alone (Control), cocultured with HCC cells (Coculture), or *PPARG* knockdown before the coculture (Coculture+sh*PPARG*) or rescue of *PPARG* expression in *PPARG* knockdown THP-1 macrophages before the coculture (Coculture+sh*PPARG*+sh*PPARG*<sub>SM</sub>) (n = 3 biological replicates). **e** Number of tubes formed by human umbilical vein endothelial cells (HUVECs) treated with different sets of THP-1 macrophages: Control, Coculture, Coculture+sh*PPARG* and Coculture+sh*PPARG*+sh*PPARG*<sub>SM</sub> (n = 3 biological replicates). The scale bars represent 20  $\mu$ m. In **a–e**, error bars indicate mean  $\pm$  sd and the statistical significances were determined by two-sided Student's *t* test. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001 and \*\*\*\**P* < 0.0001. Source data are provided as a

Source Data file.



## Supplementary Figure 11 | Intratumoral heterogeneity of the malignant hepatocytes.

Differences of pathway activities among **a** different hepatocyte/cholangiocyte clusters and **b** hepatocytes at different tissue types. Shown are  $t$  values from a linear model.

Pathways are sorted according to the variance of activities across all hepatocytes. The non-malignant, pro-tumorigenic and the pro-metastatic hepatocyte clusters are indicated by colors at the top of the heatmap. GSVA, gene set variation analysis; dn,

down; EMT, epithelial-mesenchymal transition; ROS, reactive oxygen species; UV,

ultraviolet; v1, version 1; v2, version 2. **c** Barplot of the difference of Bayesian

Information Criterion (BIC) between a one-component and a two-component

Gaussian Mixture Model. **d** Venn plot showing the overlap of malignant cells

identified based on sample tissue types, gene expression profiles and single-cell copy

number variation (CNV) profiles. **e** Distribution of cells in the 30 CNV-based clusters

identified by unsupervised  $k$ -means clustering based on the CNV profiles across

patients (left), the transcriptome-based clusters (center) and tissue types (right).

Among these 30 CNV clusters, 26 clusters exhibited significant CNVs that are

relevant to HCC, while the other four clusters did not exhibit any significant CNV,

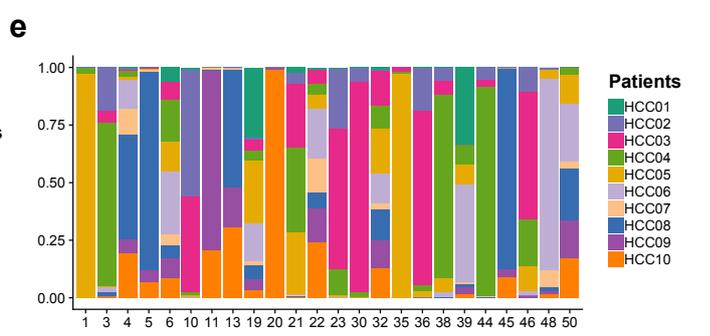
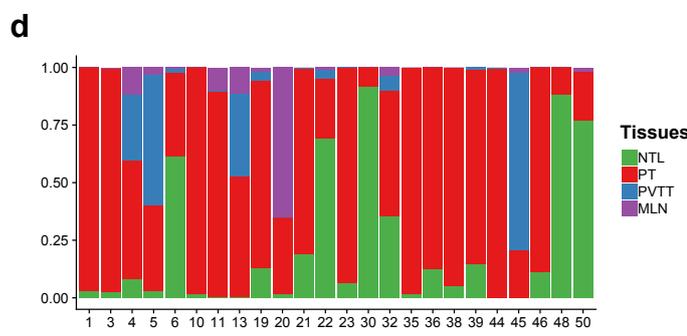
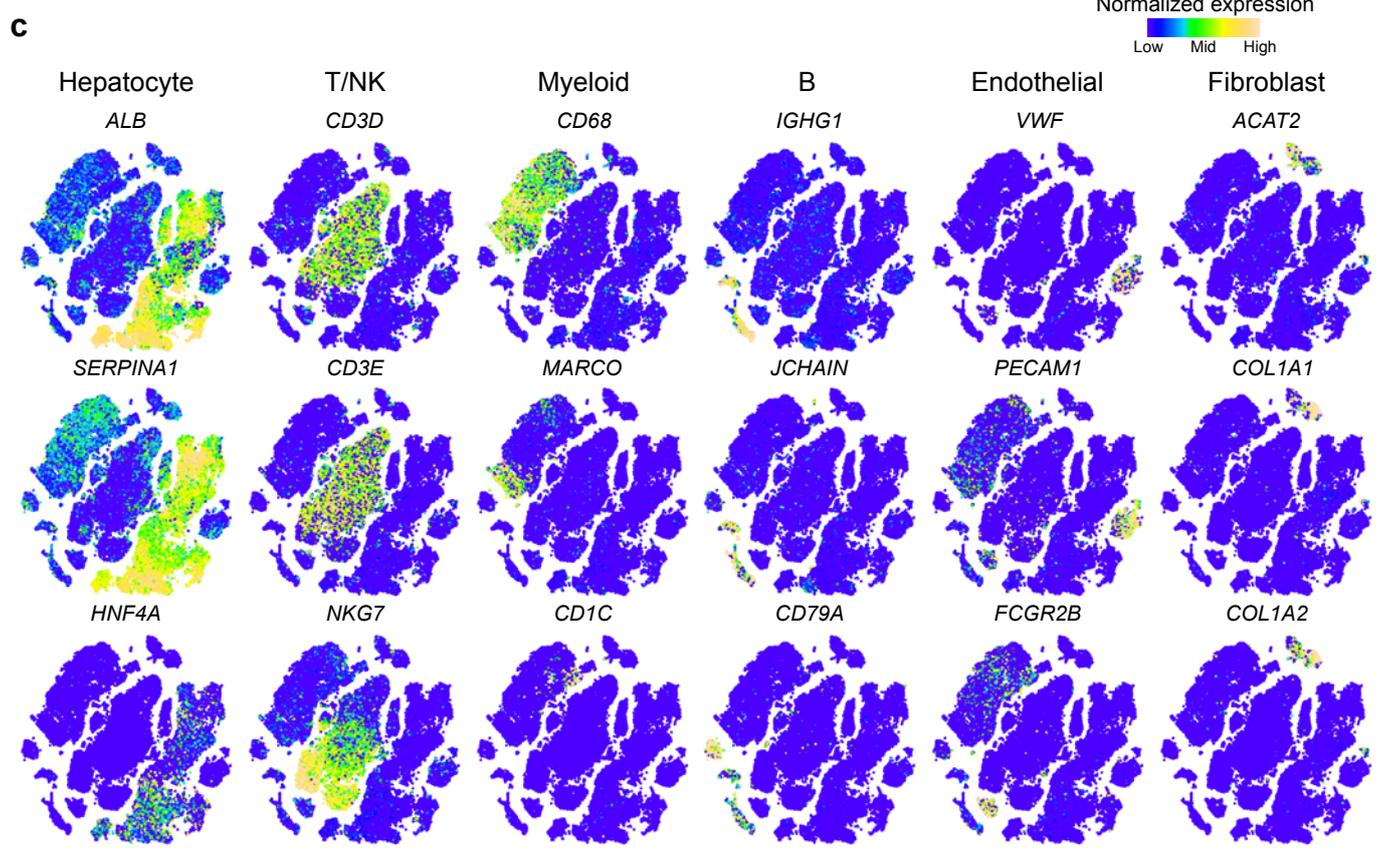
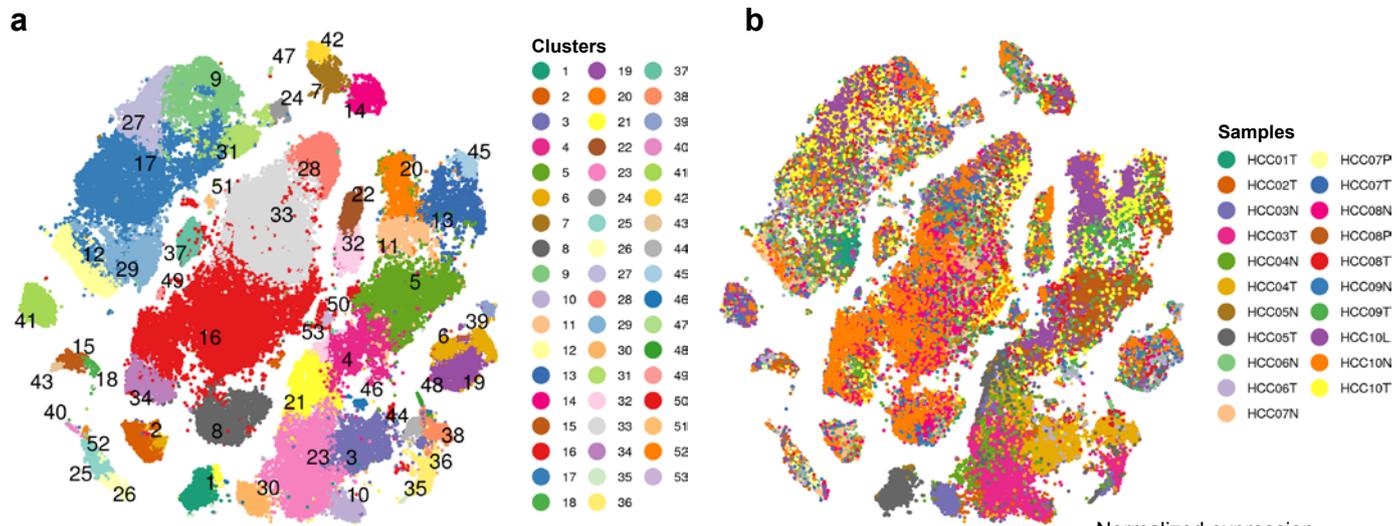
indicating they were associated to malignant and non-malignant cells, respectively.

Heatmaps (left and center) show scaled cell counts. Dot size indicates the percentage

of cells from the respective tissue types, and dot color indicates cell counts. NTL,

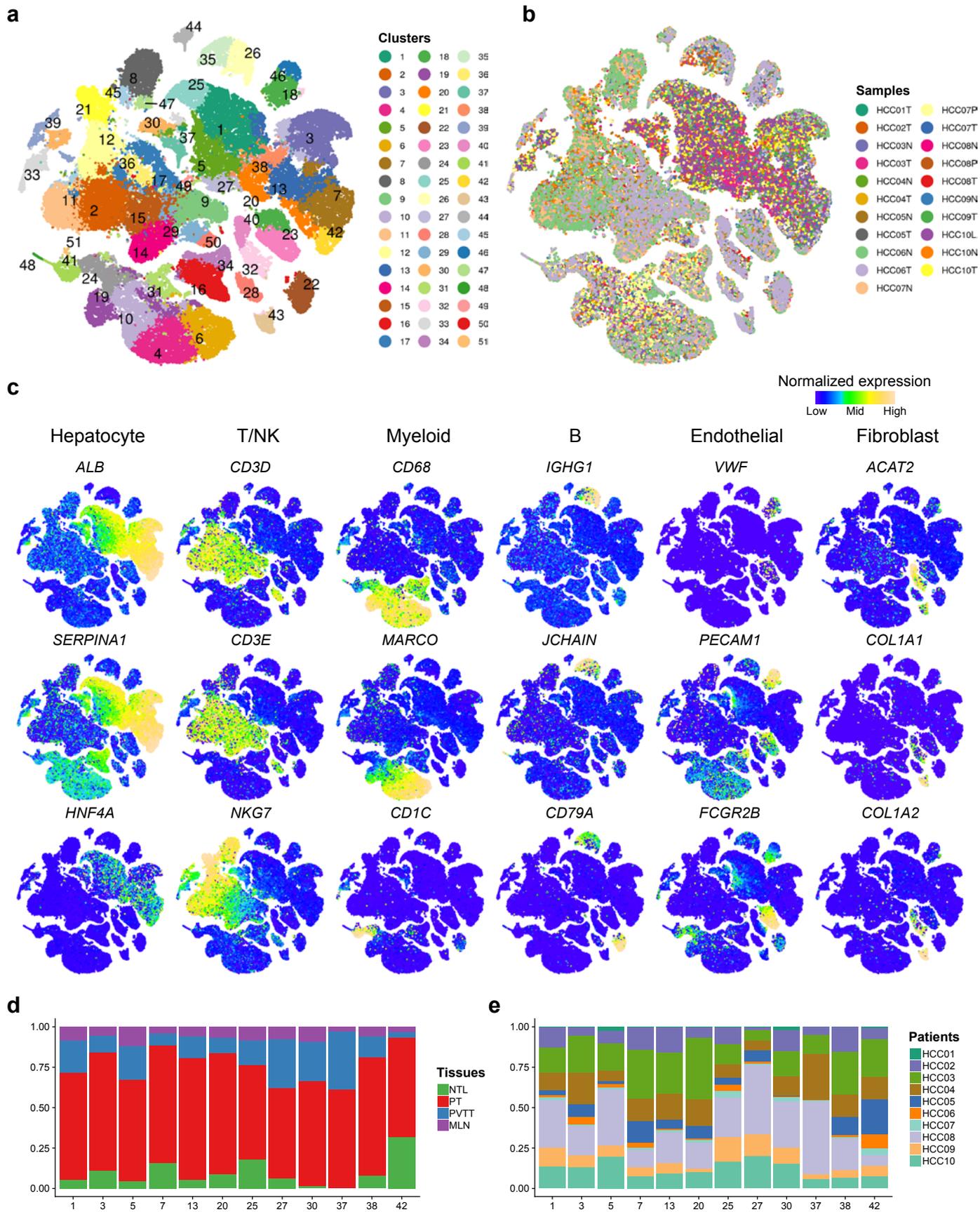
non-tumor liver; PT, primary tumor; PVTT, portal vein tumor thrombus; MLN,

metastatic lymph node. Source data are provided as a Source Data file.



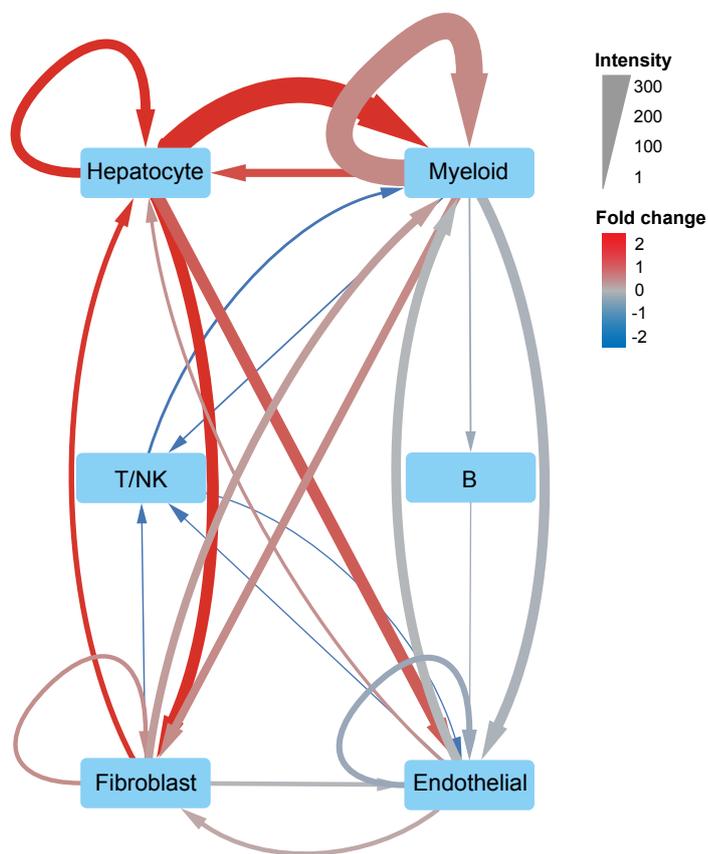
**Supplementary Figure 12 | Batch correction and cell clustering by the mutual nearest neighbor algorithm.**

The T-distributed Stochastic Neighbor Embedding (tSNE) plot of all cells with batch correction by the mutual nearest neighbor (MNN) algorithm implemented by the Scran package, colored by **a** cell clusters and **b** samples, respectively. **c** The tSNE plots showing the expression levels of signature genes of six major cell types, colored by gene expression. **d** The fractions of cells in hepatocyte clusters from different tissue types. NTL, non-tumor liver; PT, primary tumor; PVTT, portal vein tumor thrombus; MLN, metastatic lymph node. **e** The fractions of cells from different patients across hepatocyte clusters. Source data are provided as a Source Data file.



**Supplementary Figure 13 | Batch correction and cell clustering by the anchor-based algorithm.**

The T-distributed Stochastic Neighbor Embedding (tSNE) plot of all cells with batch correction by the anchor-based algorithm implemented by the Seurat v3 package, colored by **a** cell clusters and **b** samples, respectively. **c** The tSNE plots showing the expression levels of signature genes of six major cell types, colored by gene expression. **d** The fractions of cells in hepatocyte clusters from different tissue types. NTL, non-tumor liver; PT, primary tumor; PVTT, portal vein tumor thrombus; MLN, metastatic lymph node. **e** The fractions of cells from different patients across hepatocyte clusters. Source data are provided as a Source Data file.

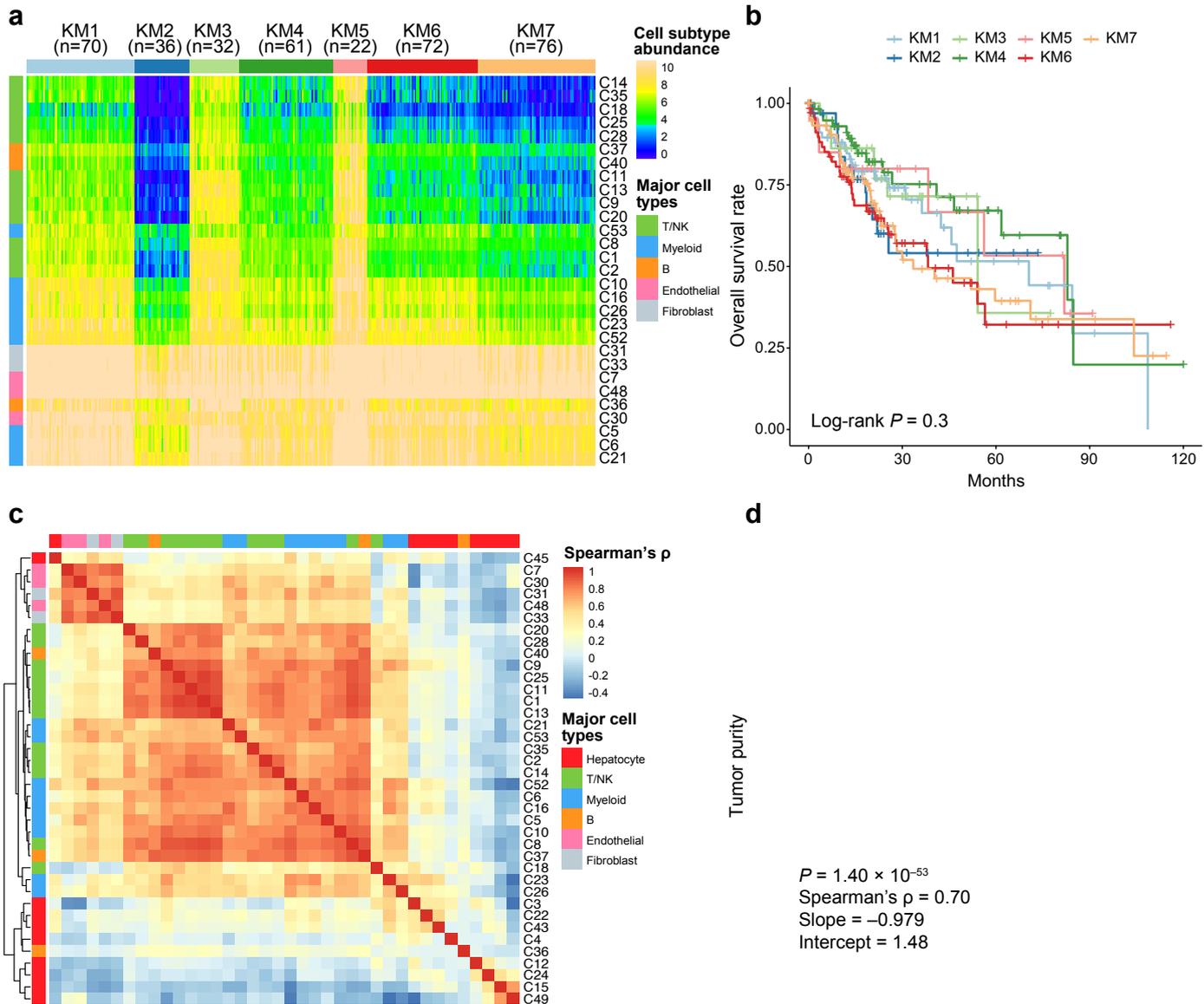
**a****b**

MYC targets v1  
 E2F targets  
 G2M checkpoint  
 MYC targets v2  
 Spermatogenesis  
 WNT/ $\beta$ -Catenin signaling  
 DNA repair  
 Unfolded protein response  
 Notch signaling  
 EMT  
 Mitotic spindle  
 Apical surface  
 Allograft rejection  
 TGF- $\beta$  signaling  
 Apical junction  
 PI3K/AKT/mTOR signaling  
 Myogenesis  
 P53 pathway  
 TNF $\alpha$  signaling via NF $\kappa$ B  
 mTORC1 signaling  
 Hedgehog signaling  
 UV response up  
 Pancreas\_ $\beta$ -Cells  
 Apoptosis  
 Hypoxia  
 Estrogen response early  
 IL2/STAT5 signaling  
 Angiogenesis  
 Glycolysis  
 Inflammatory response  
 IL6/JAK/STAT3 signaling  
 Protein secretion  
 Estrogen response late  
 UV response dn  
 ROS pathway  
 KRAS signaling dn  
 Oxidative phosphorylation  
 Androgen response  
 Heme metabolism  
 IFN- $\alpha$  response  
 Cholesterol homeostasis  
 IFN- $\gamma$  response  
 KRAS signaling up  
 Fatty acid metabolism  
 Adipogenesis  
 Complement  
 Peroxisome  
 Xenobiotic metabolism  
 Coagulation  
 Bile acid metabolism

$t$  value of GSVA score  
 (pro-metastatic versus pro-tumorigenic)

**Supplementary Figure 14 | The enriched ligand-receptor interactions among the cell types in the microenvironment of PVTT tissues.**

**a** A directed network showing the differential ligand-receptor (L-R) interaction intensities among the six major cell types (hepatocytes, T/NK, myeloid, B, endothelial cells and fibroblasts) between the portal vein tumor thrombus (PVTT) and non-tumor liver tissues. The averaged L-R interaction intensities in PVTT and non-tumor liver tissues are represented by edge width. Fold change of L-R interaction intensities is represented by edge color, with red denoting up-regulated and blue denoting down-regulated in PVTT as compared to non-tumor liver tissues. **b** Differences in pathway activities scored per cell by GSVA between pro-metastatic ( $n = 11,624$ ) and pro-tumorigenic hepatocytes ( $n = 7,285$ ). Shown are  $t$  values from a linear model. Significantly up- and down-regulated pathways (adjusted  $P$  value  $< 0.01$ ) were marked with red and blue bars, respectively, and grey bars denote non-significant pathways. dn, down; EMT, epithelial-mesenchymal transition; ROS, reactive oxygen species; UV, ultraviolet; v1, version 1; v2, version 2. Source data are provided as a Source Data file.



**Supplementary Figure 15 | Identification of the TME subtypes of HCCs using cell type deconvolution analyses.**

**a** Heatmap showing seven TME classes (KM1-7) of the HCCs identified by *k*-means clustering using the estimated abundance of non-malignant cell subtypes. Rows represent the 29 patient-shared non-malignant cell subtypes identified by the scRNA-seq data. Columns represent the 369 HCC patients in the TCGA-LIHC cohort. **b** Kaplan-Meier survival analyses showing that the seven TME classes without normalization exhibit mixed survival outcomes. The statistical significance was determined by log-rank test. **c** The correlation between the estimated abundance of 40 patient-shared malignant and non-malignant cell subtypes. Two highly correlated blocks are related to immune cells (T/NK, myeloid and B cells) and stromal cells (endothelial cells and fibroblasts) respectively. Shown are the Spearman's  $\rho$  between the estimated abundance of cell types across 369 HCC patients. Major cell types of these clusters are indicated by colors at the left and top of the heatmap. **d** Scatter plot showing the correlation between the standardized intensity of non-malignant cells and tumor purity across 369 HCC tumors from the TCGA-LIHC cohort. The solid line is the result of fitting a linear regression model. The statistical significance was determined by one-way analysis of variance (ANOVA). Source data are provided as a Source Data file.