Viral-Track integrated single-cell RNA-sequencing reveals HBV lymphotropism and immunosuppressive microenvironment in HBV-associated hepatocellular carcinoma

Zhepeng Liu[1,2,†], Wajeeha Naz[1,†], Tanzeel Yousaf[1,†], Junwei Sun[3,†], Qijia Wu[4], Mingxiong Guo[5,6,*], Geng Tian[7,*], and Guihong Sun[1,8,*]

1 Taikang Medical School (School of Basic Medical Sciences), Wuhan University, Wuhan 430071, Hubei, P.R. China

2 Department of Oncology, The First Affiliated Hospital of Wannan Medical College, Yijishan Hospital, Zaishan West Road, Wuhu 241000, Anhui, P.R. China

3 Department of Hepatic & Biliary & Pancreatic Surgery, Hubei Cancer Hospital, Affiliated Hubei Cancer Hospital of Huazhong University of Science and Technology, Wuhan 430079, Hubei, P.R. China

4 Wuhan Seqhealth Technology Co., Ltd, Wuhan 430205, Hubei, P.R. China

5 Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan 430072, Hubei, P.R. China

6 School of Ecology and Environment, Tibet University, Lhasa 850000, Xizang, P.R. China

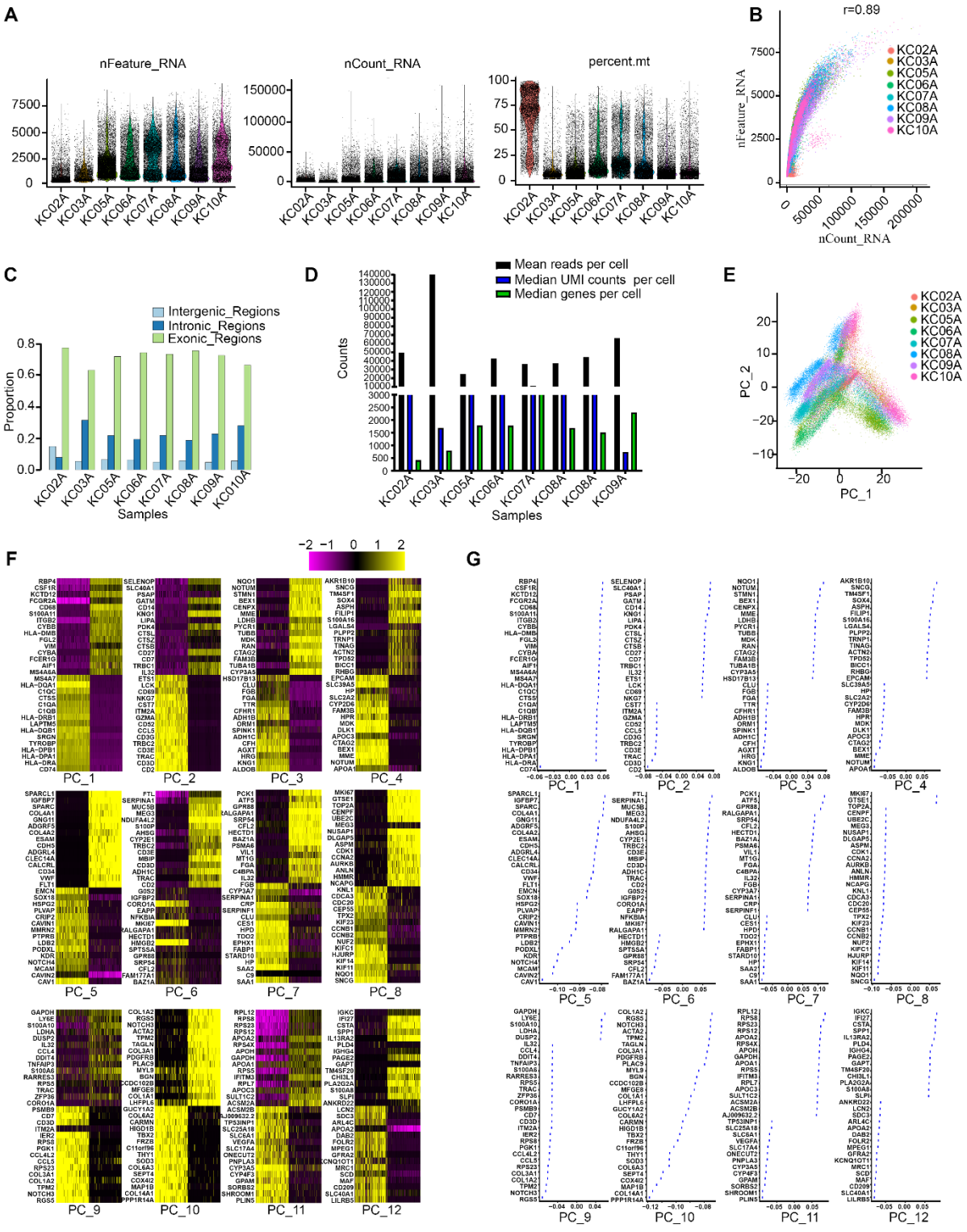
7 Department of Oncology, The First Affiliated Hospital of Shenzhen University, The Second People's Hospital of Shenzhen, Shenzhen 518035, Guangdong, P.R. China

8 Hubei Provincial Key Laboratory of Allergy and Immunology, Wuhan 430071, Hubei, P.R. China

* Corresponding author. Addresses: Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan 430072, Hubei, P.R. China (Mingxiong Guo); or Department of Oncology, The First Affiliated Hospital of Shenzhen University, The Second People's Hospital of Shenzhen, Shenzhen 518035, Guangdong, P.R. China (Geng Tian); or Taikang Medical School (School of Basic Medical Sciences), Wuhan University, Wuhan 430071, Hubei, P.R. China (Guihong Sun)

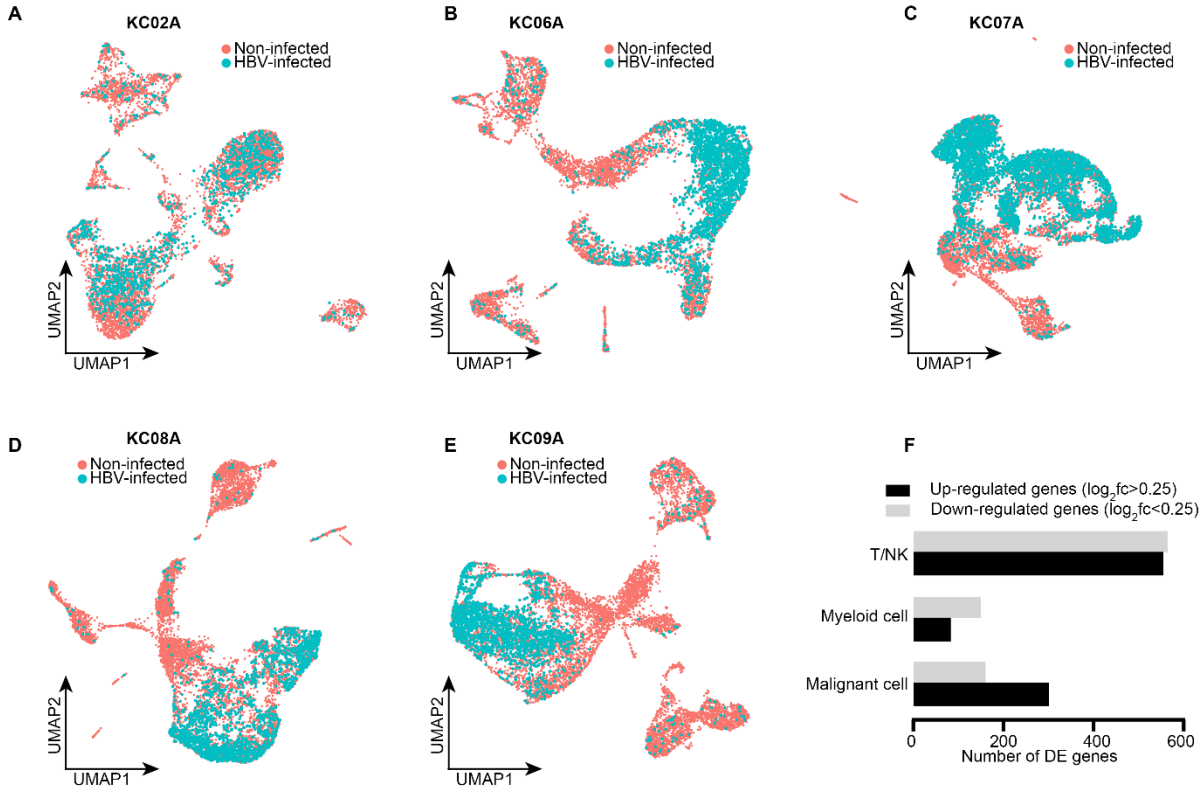
E-mail addresses: guomx@whu.edu.cn (Mingxiong Guo), tiangeng_tg666@163.com Geng Tian), ghsunlab@whu.edu.cn (Guihong Sun).

† Z.L., W.N., T.Y., and J.S. These authors contributed equally.

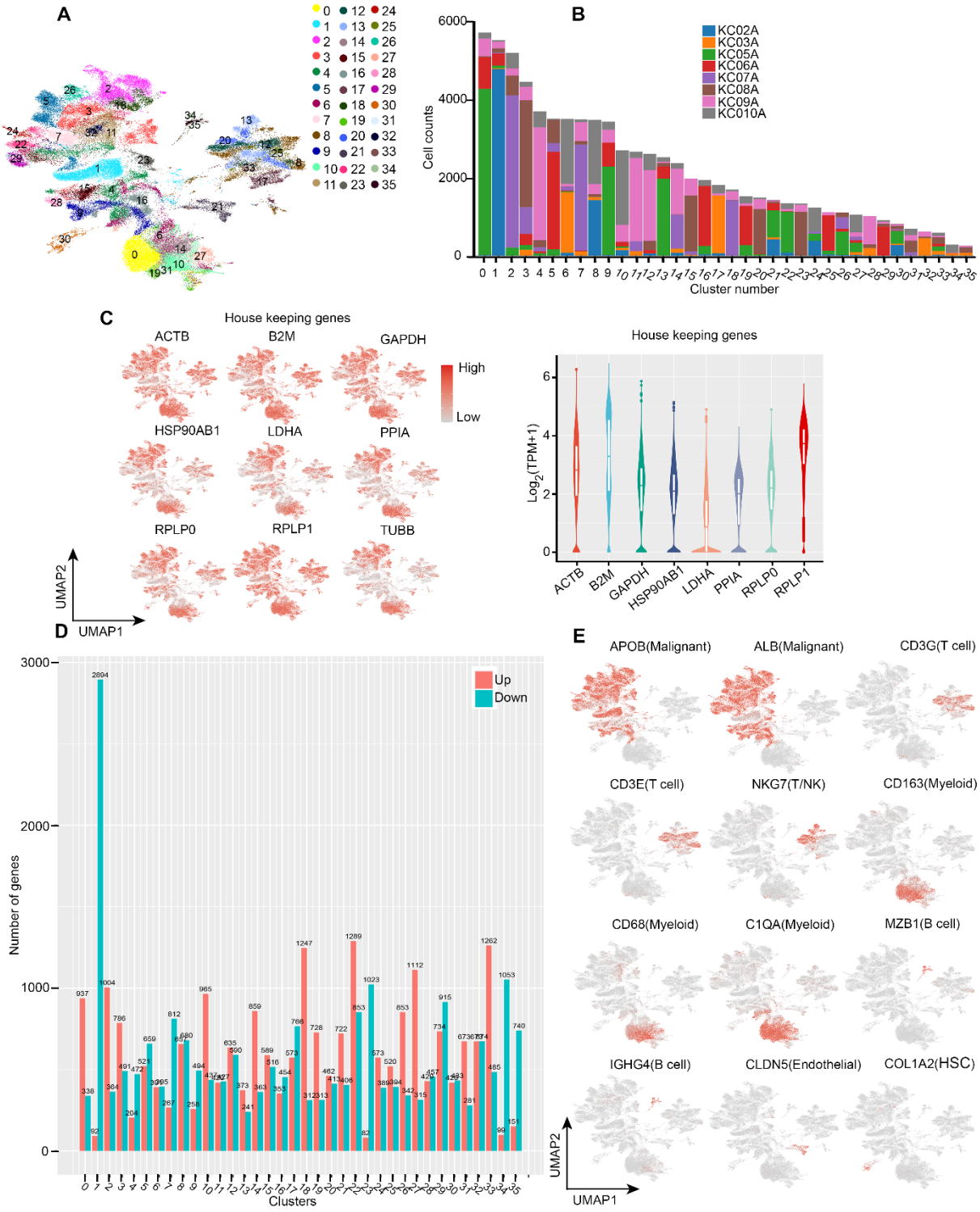


Supplementary Figure 1. Quality Control Analysis of scRNA-Sequence data, related to Figure 1. **A.** Violin diagram representation of quality control checks: nFeature_RNA, nCount_RNA, and percent. Mt in each cell in the sample. **B.** Correlation plot between the number of gene expressions and the number of unique UMIs (Sequencing depth), color-coded, according to patient. The detected gene

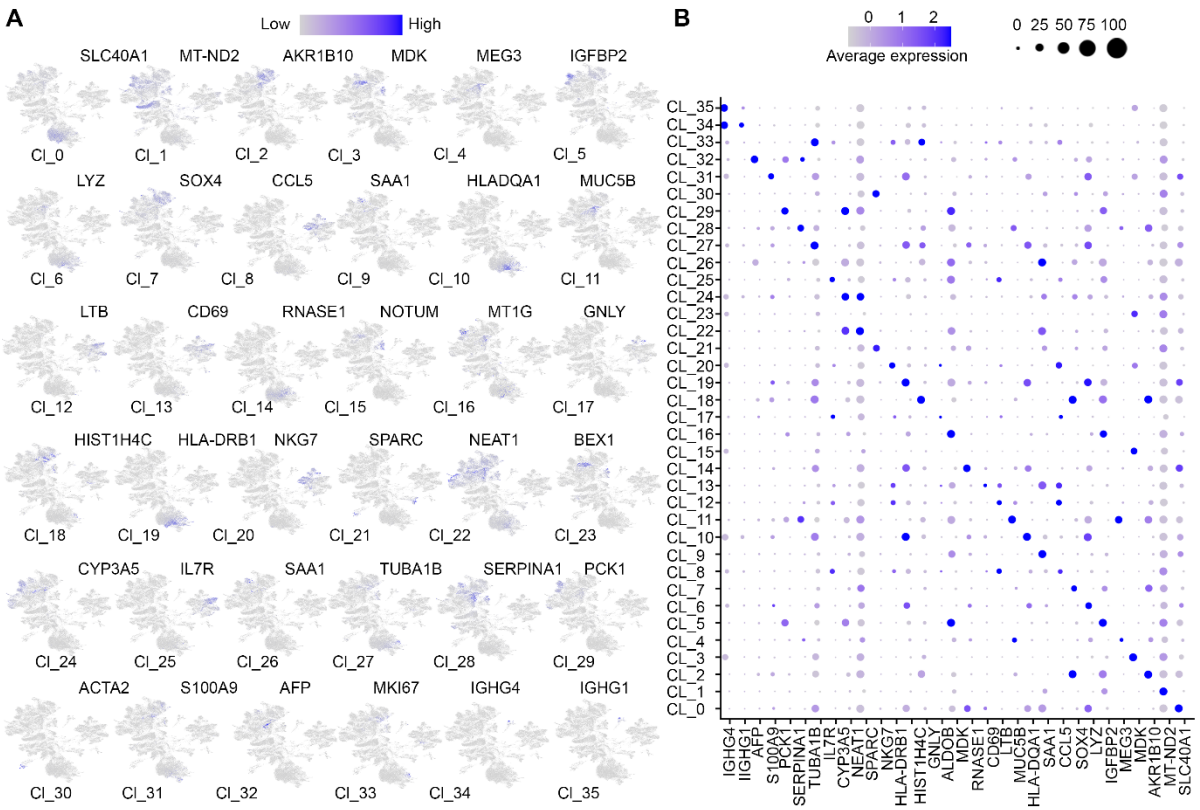
numbers positively correlated with sequencing depth. **C.** Histograms of sequencing reads data indicate the proportions of exonic, intronic, and intergenic genomic regions in each sample. **D.** Histograms indicate the number of genes detected per cell and the summary of total mapped reads. **E.** Scatter plot of the first two principal components, color-coded, according to patient. **F&G.** Heatmap and dot plot of the top 12 principal component genes expression. PCA was conducted to identify the significantly available dimensions of the data sets with estimated P values. Cells were classified by PCA.



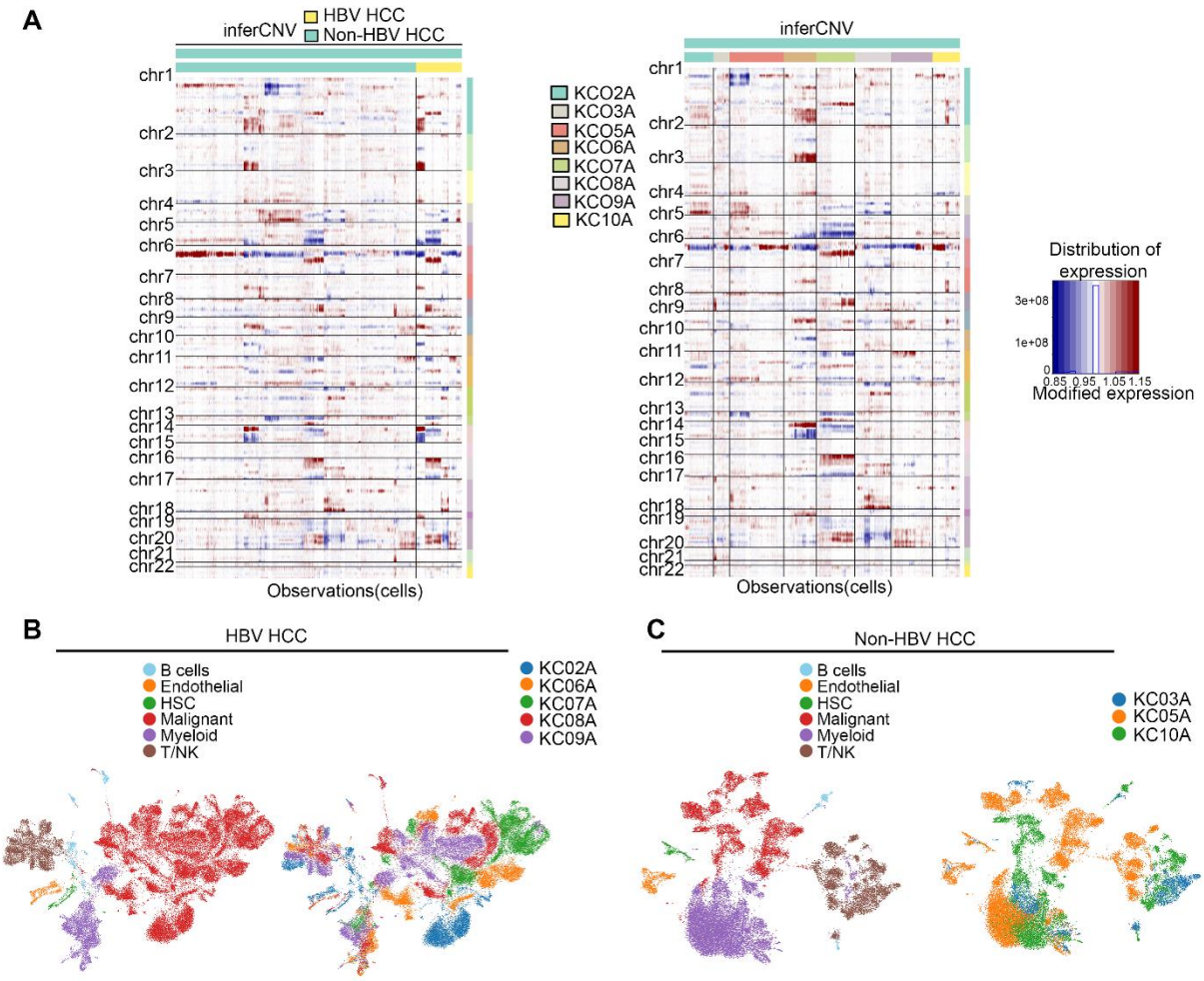
Supplementary Figure 2. Viral-track analysis of HBV HCC patients to determine HBV-infected and non-infected cells, related to Figure 1. A-E. UMAP results of Viral-Track analysis showing infected (green) and non-infected (red) cells inferred from scRNA-seq data, and determined the HBV HCC patients, (A) KC02A, (B) KC06A, (C) KC07A, (D) KC08A, and (E) KC09A. **F.** Number of differentially expressed genes between non-infected and infected T/NK cells, myeloid, and malignant cells.



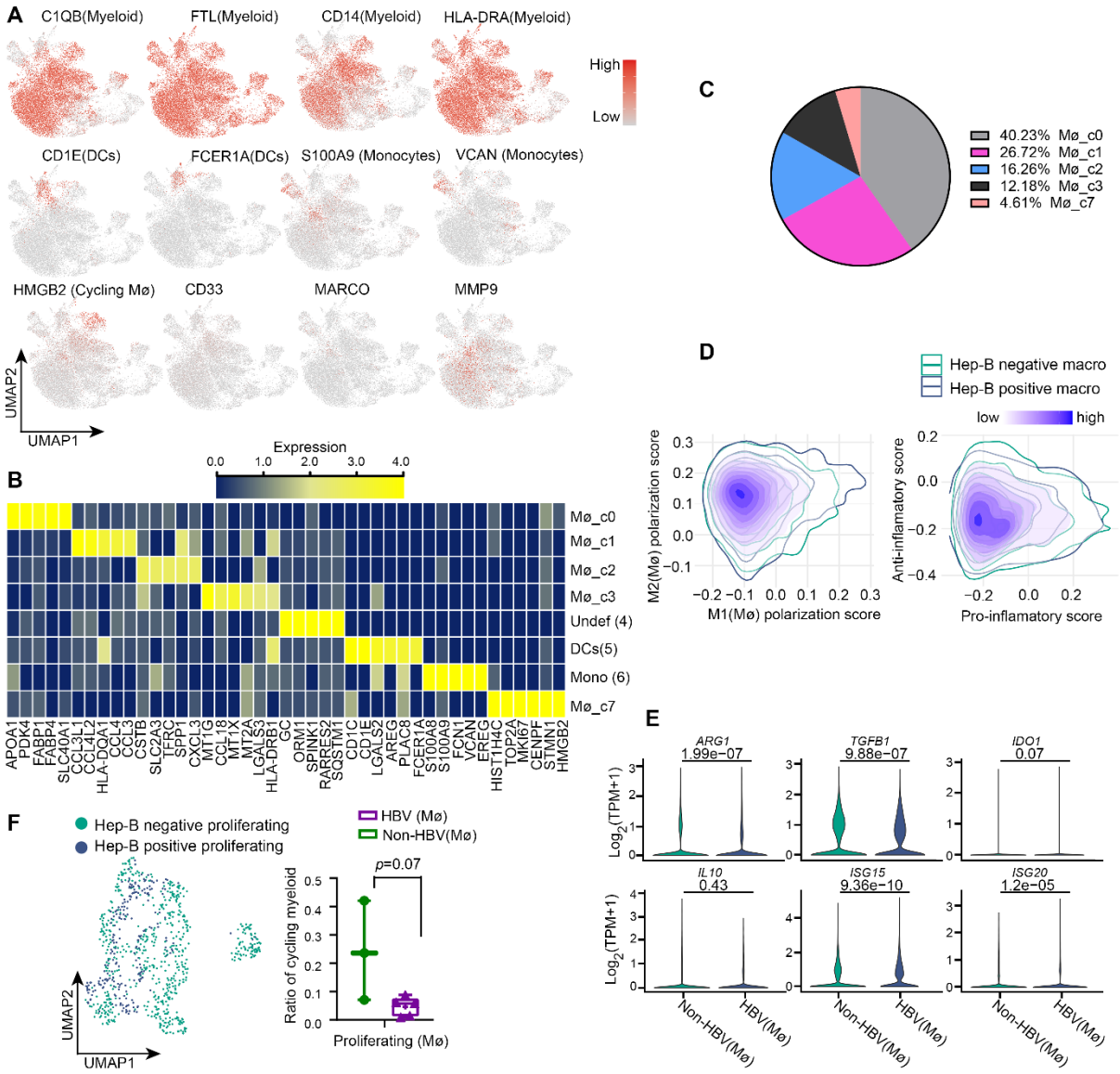
Supplementary Figure 3. The landscape of the tumor ecosystem in HBV and non-HBV-HCC tumor tissues is related to Figure 2. A. UMAP plot, showing the cluster origins by color and annotation for cluster number in the HCC ecosystem. **B.** The cell count representation in all 36 clusters of each analyzed patient. **C.** UMAP plots (left panel) and violin plots (right panel) showing the expression levels of 9 known housekeeping genes defined in all cells. **D.** The bar graph indicating the number of up-regulated and down-regulated genes in all 36 clusters. **E.** Expression of marker genes correlated with immune and non-immune cell types.



Supplementary Figure 4. The landscape of the tumor ecosystem in HBV and non-HBV-HCC tumor tissues is related to Figure 2. A-B. UMAP (A) and dot plots (B) showing the highly expressed marker genes for all clusters (0-35), defined for all cell types.

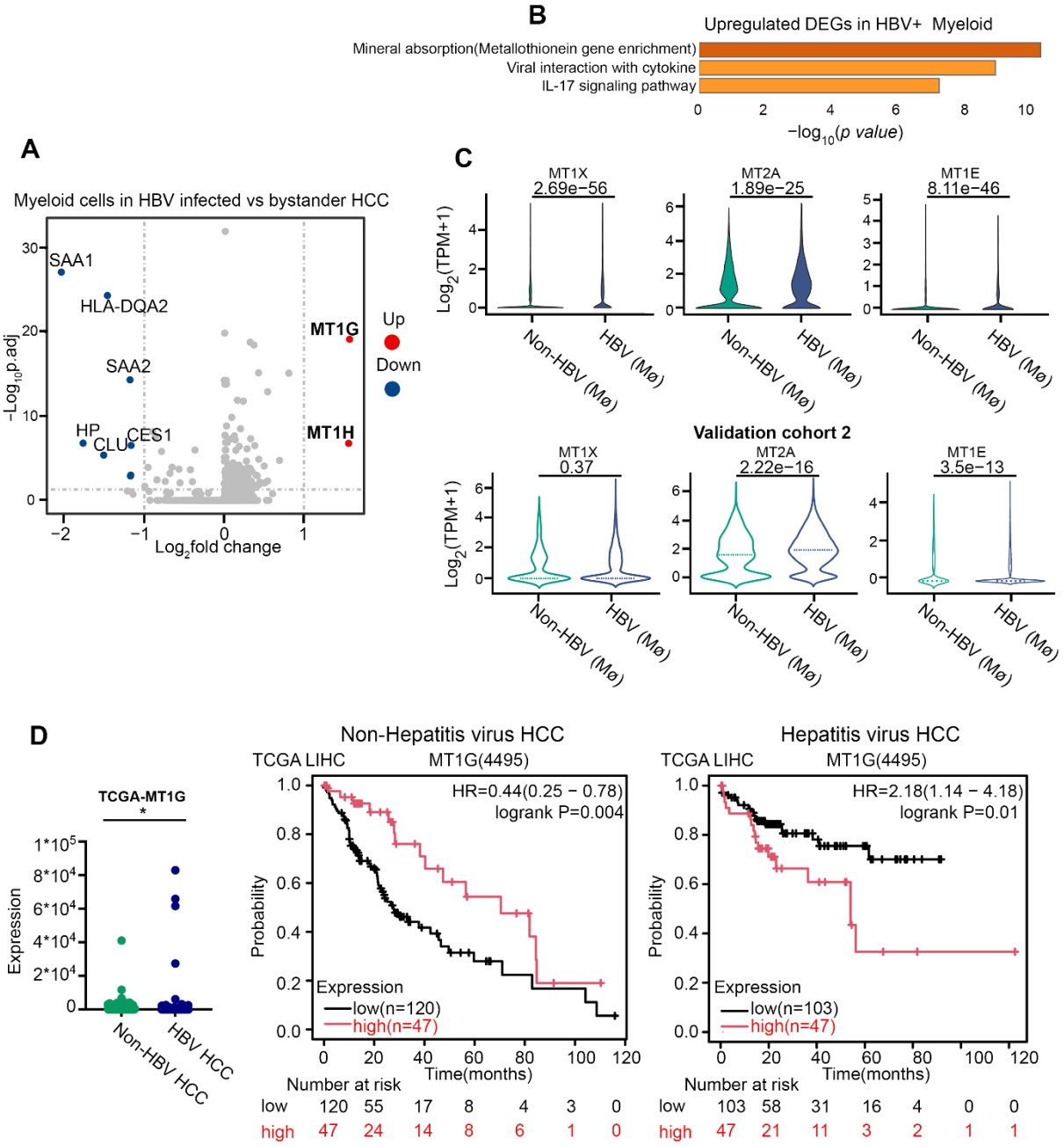


Supplementary Figure 5. The landscape of the tumor ecosystem in HBV- and non-HBV-HCC tumor tissues is related to Figure 2. A. Heatmap representing the CNV analysis, inferred from the single-cell RNA-seq data for each tumor cell. Each column represents tumor type (left panel), and each patient analyzed (right panel) is color-coded. **B.** The UMAP plot showing color codes for cell types (left panel) and patient origin (right panel) in the HBV HCC ecosystem. **C.** The UMAP plot showing color codes for cell types (left panel) and patient origin (right panel) in the non-HBV HCC ecosystem.



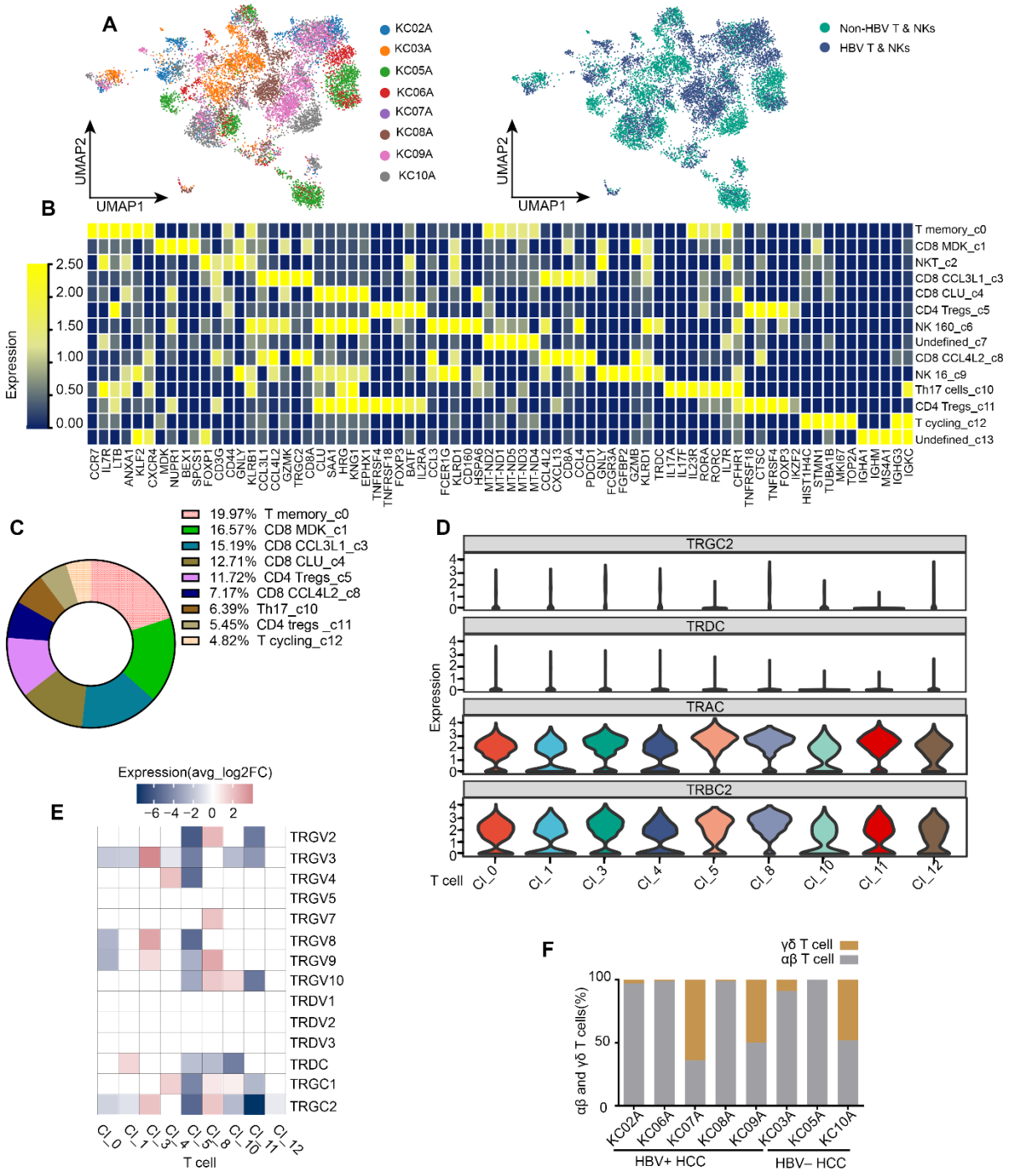
Supplementary Figure 6. The prevalence of metallothioneins-enriched TAMs in the HBV-mediated HCC microenvironment is related to Figure 3. A. Expression of selected marker genes correlated with myeloid subtypes (macrophages, monocytes, and DCs). **B.** A scaled heatmap showing the expression of marker genes in each subtype of myeloid-derived cells. **C.** Pie chart showing the percentage distribution of each macrophage cluster. **D.** Scatter plots showing M1 and M2 polarization scores (left panel) and pro- and anti-inflammatory scores (right panel), color-coded according to HBV and non-HBV tumor type. **E.** Violin plots indicating the expression of selected genes (ARG1, TGFB1, IDO1,1L10, ISG15, and ISG20) from macrophage clusters across HBV and non-HBV tumors. **F.** UMAP visualization of reclustered cycling macrophages, showing HBV and non-HBV tumor type by color (left panel). Boxplot showing the fractions of cycling myeloid in HBV (purple) and non-HBV HCC (green) (right panel). Error bars represent mean ± standard deviation (SD). The p-values in (E) were calculated using the Student's t-test, and in (F), were calculated by using the unpaired Wilcoxon

test. All statistical comparisons were performed on biologically independent samples (n = 5 HBV-HCC and n = 3 non-HBV-HCC) from the discovery cohort.



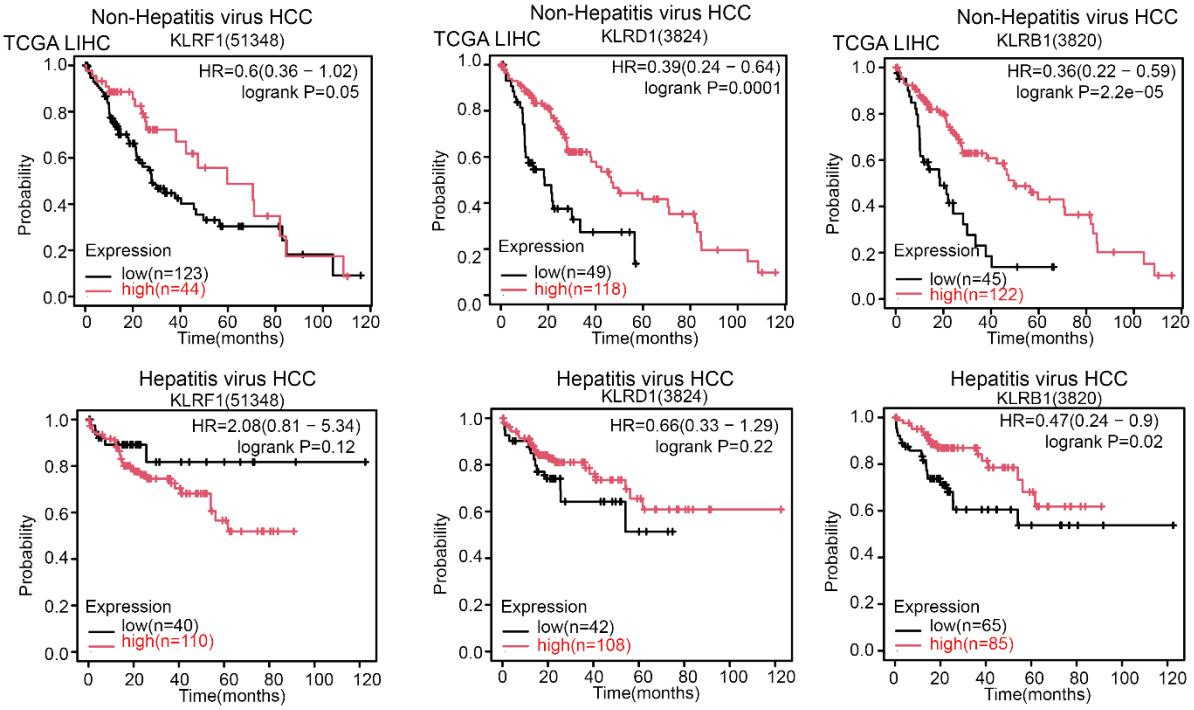
Supplementary Figure 7. The prevalence of metallothioneins-enriched TAMs in the HBV-mediated HCC microenvironment is related to Figure 3. A. The volcano plot showing differentially expressed genes between infected (red dots) and non-infected myeloid (blue dots). The names of the most significant genes are indicated in the plots. **B.** Bar chart showing the enrichment of KEGG pathways based on the upregulated genes from HBV HCC macrophages. **C.** Violin plots indicating the expression of metallothionein genes (MT1X, MT2A, and MT1E) from macrophage clusters between HBV (blue) and non-HBV (green) tumor (top panel), violin plots depicting the MT1X, MT2A, and

MT1E genes expression in HBV (blue) and non-HBV (green) samples from the Validation cohort 2 (bottom panel). **D.** Expression of MT1G in HBV (blue) and non-HBV (green) from the TCGA-LIHC database by OncoDB (left panel). The Kaplan-Meier plot of disease-free survival in the publicly available TCGA LIHC cohort, comparing low versus high MT1G in non-hepatitis virus (middle panel) and Hepatitis virus HCC patients (right panel). The number of patients from each subgroup and the Log-Rank t-test results are indicated in the plot. N represents the number of patients. The p-values were calculated using the Student's t-test in (C) &(D). All statistical comparisons were performed on biologically independent samples (n = 5 HBV-HCC and n = 3 non-HBV-HCC) from both the discovery and validation cohort 2 in (C).

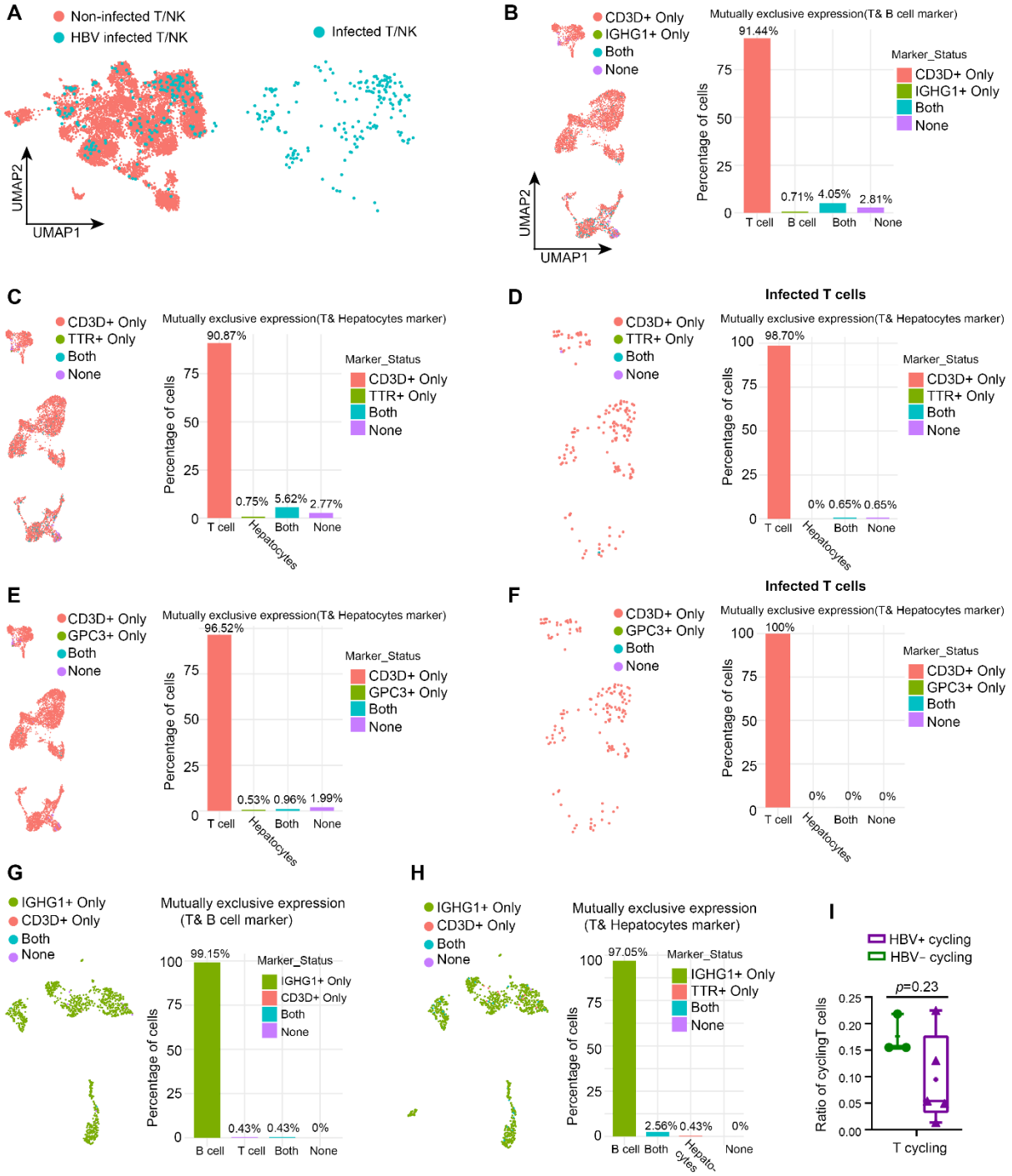


Supplementary Figure 8. Investigating T/NK cells functional diversity and HBV lymphotropism in HBV-mediated HCC, related to Figure 4. A. The UMAP plot showing patient IDs by color (left panel); the UMAP plot showing tumor type by color, HBV, or non-HBV HCC (right panel). **B.** Heatmap showing the scaled, average expression of typical marker genes for each cluster defined in Figure 5A. **C.** Pie chart showing the percentage distribution of each T cell cluster. **D.** Violin plots displaying marker gene expression from γδ T cells (TRGC2 and TRDC) or αβ T cells (TRAC and TRBC2) in each cluster. **E.** Heatmap displaying T cell receptor gamma and delta variable region genes (TRGV and TRDV) or T cell receptor gamma and delta constant region genes (TRGC and TRDC) expression patterns in T cell

clusters. **F.** Histogram indicating the percentage (%) distribution of γδ and αβ T cells in tumor tissue of each analyzed patient.

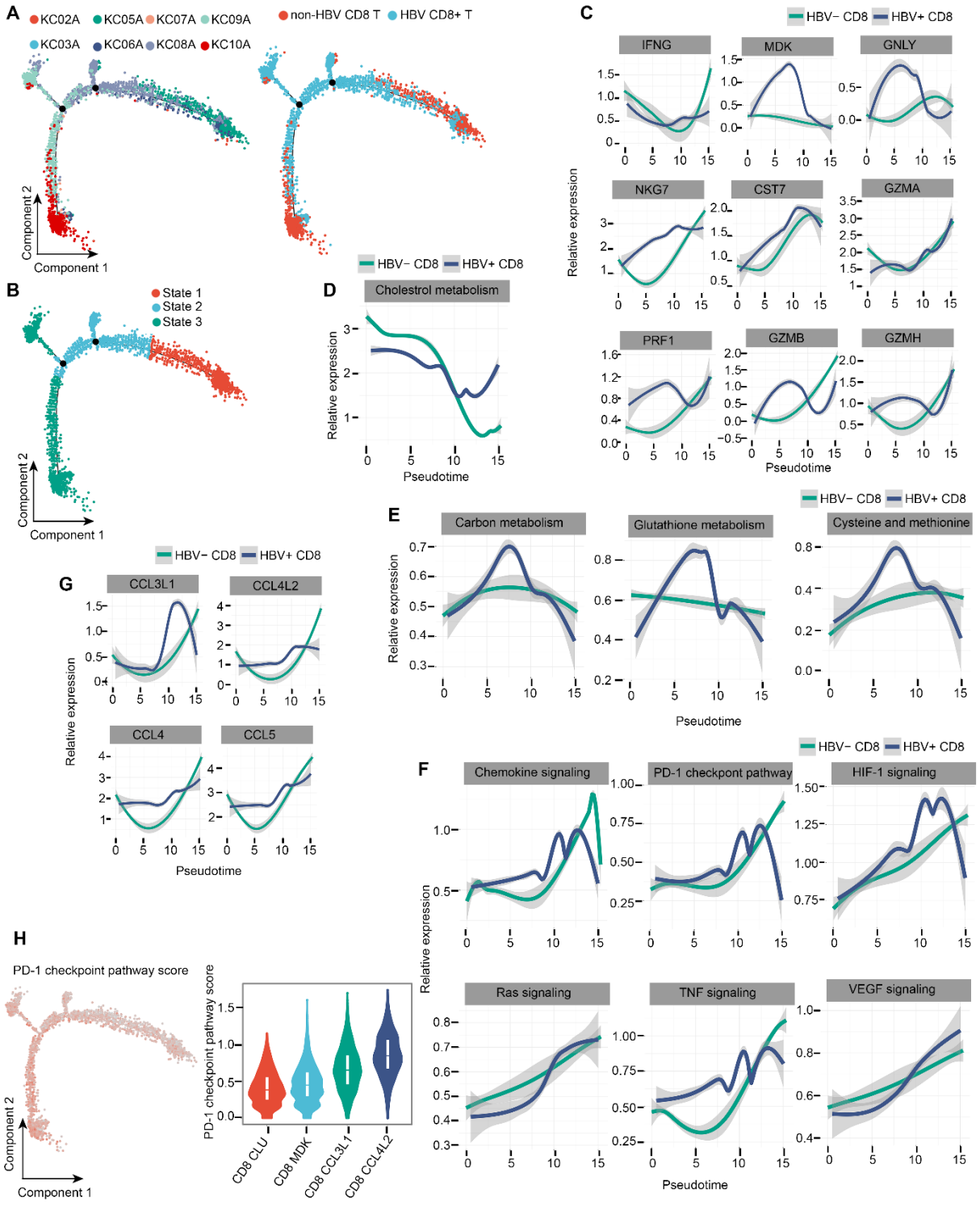


Supplementary Figure 9. Investigating T/NK cells functional diversity and HBV lymphotropism in HBV-mediated HCC, related to Figure 4. Kaplan-Meier plots of disease-free survival in the publicly available TCGA LIHC cohort, comparing low versus high KLRF1 (left panel), KLRD1 (middle panel), and KLRB1 (right panel) in Hepatitis virus (bottom panel) and non-Hepatitis (top panel) virus HCC patients. The numbers of patients from each subgroup and the Log-Rank t-test results are indicated in the plot. N represents the number of patients.



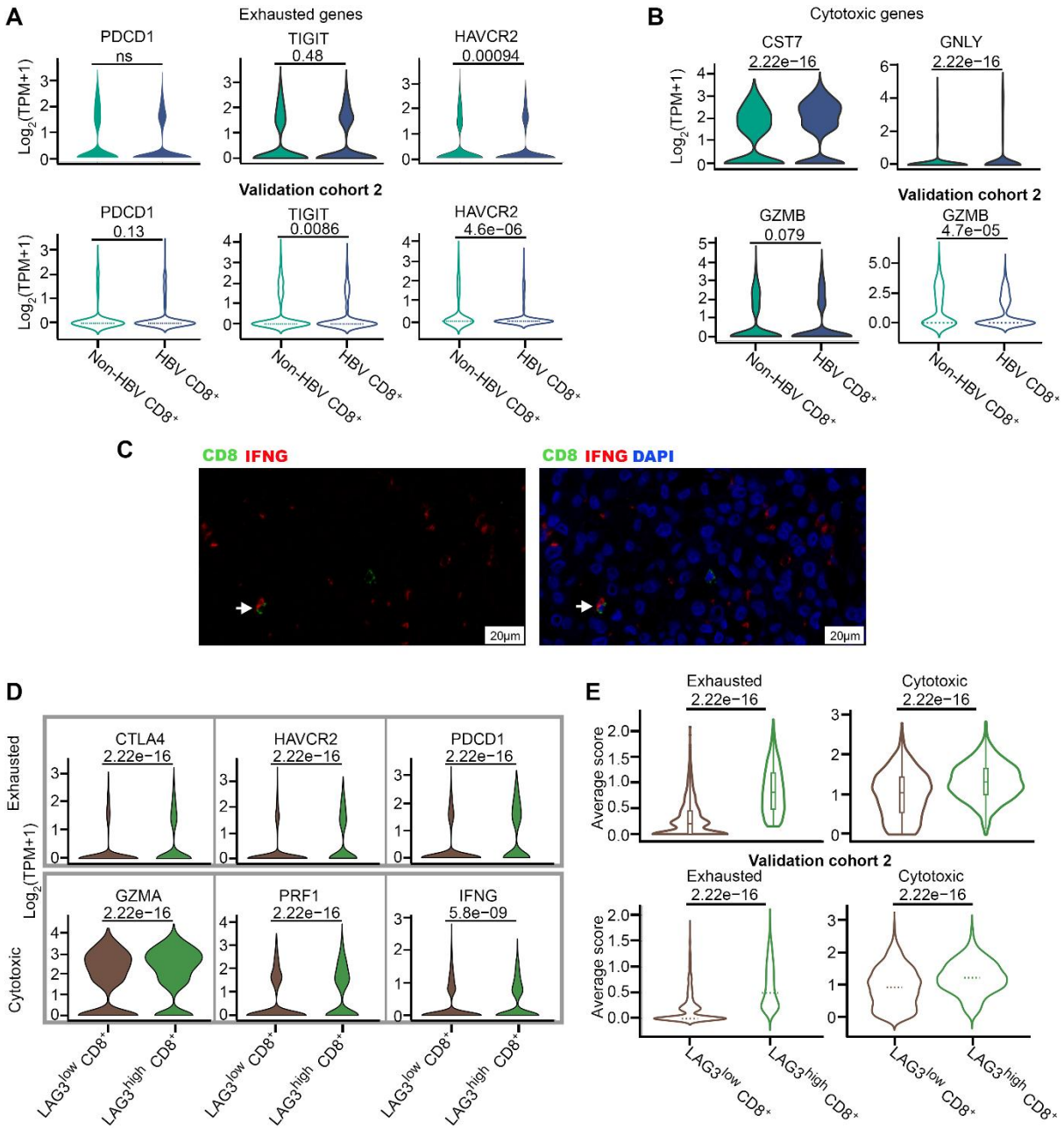
Supplementary Figure 10. Investigating T/NK cells functional diversity and HBV lymphotropism in HBV-mediated HCC, related to Figure 4. A. UMAP colored by viral status, representing HBV-infected (green) and non-infected (red) T/NK cells. **B.** UMAP plot illustrates the mutually exclusive expression of only T cell ($CD3D^+$, red), only B cell ($IGHG1^+$, green), both markers (blue), and none (purple) markers expression, showing distinct expression profiles in HBV-HCC patients. **C**. UMAP plot illustrates the mutually exclusive expression of only T cell ($CD3D^+$, red), only malignant hepatocytes (TTR^+, green), both markers (blue) and none (purple) markers expression showing distinct expression profiles in all T cells from HBV-HCC patients. **D**. UMAP plot illustrates the mutually exclusive expression of only T cell ($CD3D^+$, red), only malignant hepatocytes (TTR^+, green), both

markers (blue) and none (purple) markers expression showing distinct expression profiles in HBV-infected T cells. **E**. UMAP plot illustrates the mutually exclusive expression of only T cell ($CD3D^+$, red), only malignant hepatocytes ($GPC3^+$, green), both markers (blue) and none (purple) markers expression showing distinct expression profiles in all T cells from HBV-HCC patients. **F**. UMAP plot illustrates the mutually exclusive expression of only T cell ($CD3D^+$, red), only malignant hepatocytes ($GPC3^+$, green), both markers (blue) and none (purple) markers expression showing distinct expression profiles in HBV-infected T cells. **G.** The UMAP plot illustrates the mutually exclusive expression of only B cells (IGHG1+, green), only T cells (CD3D+, red), both markers (blue), and none (purple) markers, showing distinct expression profiles in all B cells from HBV-HCC patients. **H**. The UMAP plot illustrates the mutually exclusive expression of only B cells (IGHG1+, green), only malignant hepatocytes (TTR+, green), both markers (blue), and none (purple) markers, showing distinct expression profiles in all B cells from HBV-HCC patients. **I.** Boxplot showing the fractions of cycling T cells in HBV (purple) and non-HBV-HCC (green). Error bars represent mean ± standard deviation (SD). The p-values in (I) were calculated using the unpaired Wilcoxon test. All statistical comparisons were performed on biologically independent samples (n = 5 HBV-HCC and n = 3 non-HBV-HCC) from the discovery cohort.



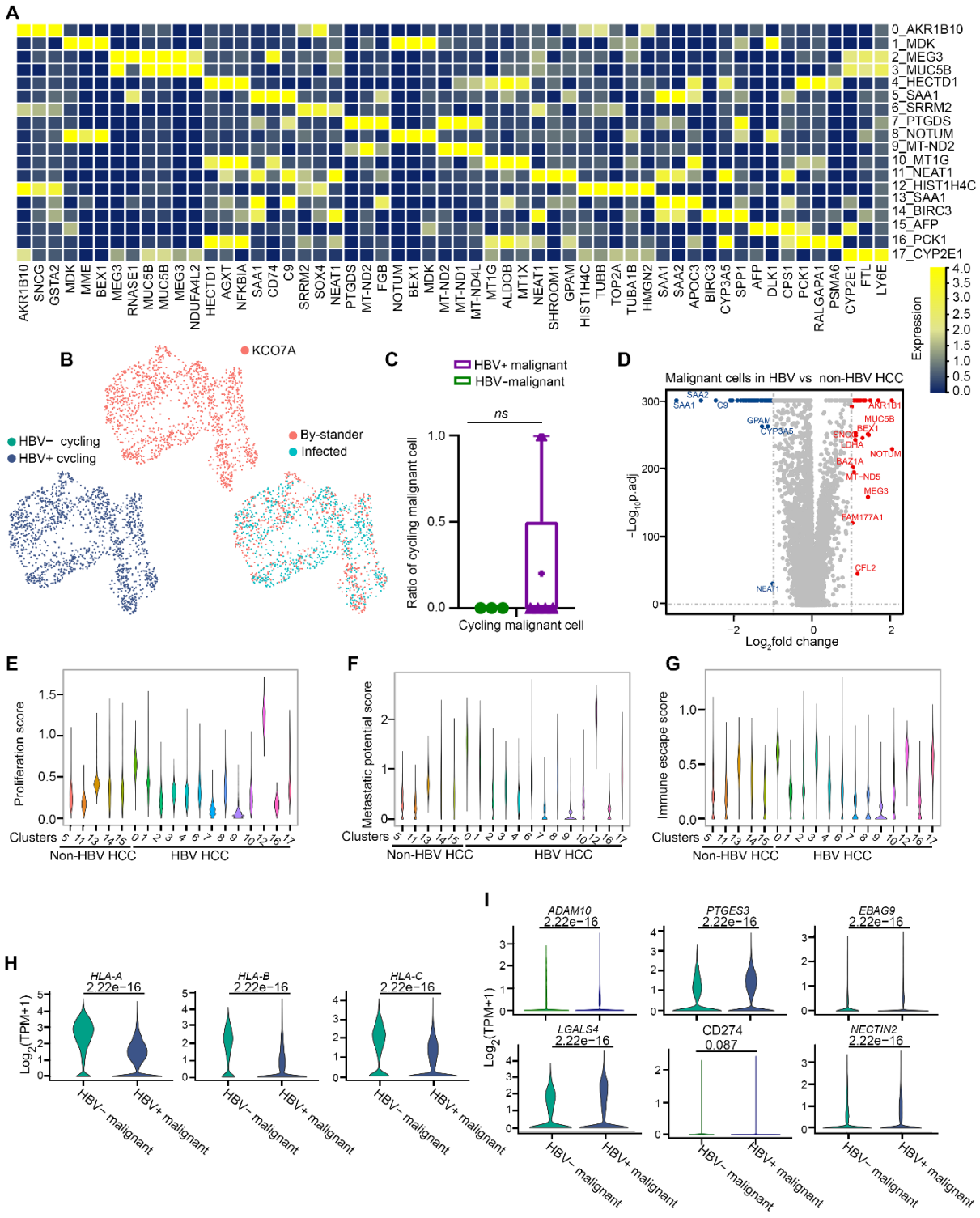
Supplementary Figure 11. Cell-state transition trajectory analysis of CD8$^+$ tumor-infiltrating T cell populations in HBV and non-HBV-HCC, related to Figure 5. A. Pseudotime-ordered analysis of CD8$^+$ T cells from non-HBV and HBV tumor samples. Patient IDs (left panel) and tumor type (right panel) are labeled by colors. **B.** Pseudotime-ordered analysis of CD8$^+$ T cells. CD8$^+$ T cell states (three) are labeled by colors. **C.** Two-dimensional plots showing the dynamic expression of MDK and cytotoxic marker genes (IFNG, NKG7, GNLY, PRF1, CST7, GZMA, GZMB, and GZMH) during the T cell transitions along the pseudotime. **D.** Two-dimensional plots showing the expression of average

scores for genes related to cholesterol metabolism in non-HBV (green) and HBV (blue) HCC samples, along with the pseudotime. **E.** Two-dimensional plots showing the expression of average scores for genes related to carbon, glutathione cysteine, and methionine metabolism in non-HBV (green) and HBV (blue) HCC samples, along with the pseudotime. **F.** Two-dimensional plots showing the expression of average scores for genes related to signaling pathways (chemokine, PD-1, HIF-1, Ras, TNF, and VEGF) in non-HBV (green) and HBV (blue) HCC samples, along with the pseudotime. **G.** Two-dimensional plots showing the dynamic expression of chemokines (CCL3L1, CCL4L2, CCL4, and CCL5) during the T cell transitions along the pseudotime. **H.** 2D pseudotime plot showing the dynamics of PD-1 checkpoint score (left panel) or violin plots (right panel) showing the average expression of PD-1 checkpoint genes from the clusters of $CD8^+$ T cells in HBV and non-HBV tumor samples.



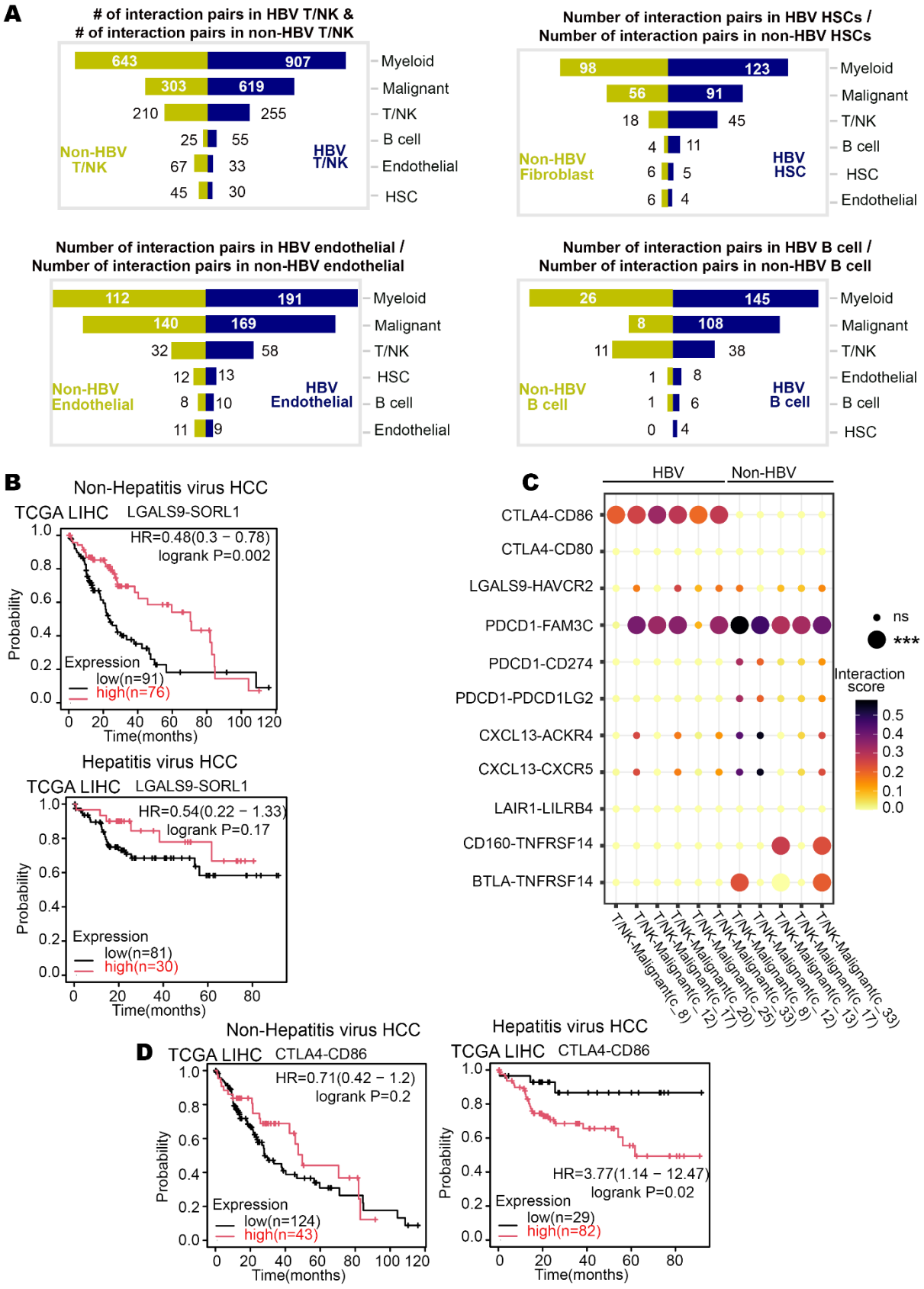
Supplementary Figure 12. HBV-related HCC tumor infiltrating MDK^{high}IFN-γ^{low}CD8^{+} T cells displayed a dysfunctional immunosuppressive phenotype, related to Figure 6. A. Violin plots indicating the expression of exhausted marker genes (PDCD1, TIGIT, and HAVCR2) in CD8^{+} T cells across HBV (blue) and non-HBV (green) tumor samples in the discovery cohort (upper panel) and validation cohort 2 (lower panel). **B.** Violin plots indicating the expression of cytotoxic marker genes (CST7, GNLY, and GZMB) in CD8^{+} T cells across HBV (blue) and non-HBV (green) tumor samples (left panel) or violin plot showing the expression of cytotoxic marker gene GZMB in HBV (blue) and non-HBV (green) samples from the Validation cohort 2 (lower panel). **C.** Representative IHC staining, showing CD8^{+}IFNG^{+} T cells in non-tumor liver tissue. Scale bar, 20μm. **D.** Violin plots indicating the expression of exhausted(CTLA4, HAVCR2, and PDCD1, upper panel) and cytotoxic marker genes (GZMA, PRF1, and IFNG, lower panel) in LAG3^{high} (log_2 expression>0.585) and LAG3^{low} (log_2

expression<0.585) $CD8^+$ T cells across HBV and non-HBV tumor samples. **E.** Violin plots indicating the average score of exhausted and cytotoxic signals in $LAG3^{high}$ (log_2 expression>0.585) and $LAG3^{low}$ (log_2 expression<0.585) $CD8^+$ T cells across HBV and non-HBV tumor samples in the discovery cohort (upper panel) and validation cohort 2 (lower panel). The p-values in (A, B, D, and E) were calculated using Student's t-test. The p-values in (H and I) were calculated using Student's t-test, and in (F) by the unpaired Wilcoxon test. All statistical comparisons were performed on biologically independent samples (n = 5 HBV-HCC and n = 3 non-HBV-HCC) from both the discovery and validation cohort 2.



Supplementary Figure 13. Highly proliferative malignant hepatocytes in HBV-related HCC foster more immunosuppressive TME than non-HBV-HCC, related to Figure 7. A. Heatmap showing the scaled, average expression of typical marker genes in malignant cells for each of the clusters defined in Figure 7A. **B.** UMAP visualization of cycling malignant cells, color-coded by patient IDs (Top panel), showing HBV and non-HBV tumor type (left panel) or infected or non-infected cells (right panel). **C.** Boxplot showing the fractions of cycling malignant cells in HBV (blue) and non-HBV tumour patients (green). Error bars represent mean ± standard deviation (SD). **D.** The volcano plot

shows differentially expressed genes between HBV^+ (red dots) and HBV^- tumor malignant cells (blue dots). The names of the most significant genes are indicated in the plots. **E-G.** Violin plots showing the expression of Proliferation(D), metastatic potential (E), and immune escape genes (F) average score from each cluster of malignant cells. **H.** Violin plots showing the expression of selected immune surveillance genes (HLA-, HLA-B, and HLA-C) in tumor cells from HBV (blue) and non-HBV (green) tumor samples. **I.** Violin plot showing the expression of selected immune escape genes (ADAM10, PTGES3, EBAG9, LGALS4, CD274, and NECTIN2) in tumor cells from HBV (blue) and non-HBV (green) tumor samples. The p-values in (H and I) were calculated using Student's t-test, and in (C) by the unpaired Wilcoxon test. All statistical comparisons were performed on biologically independent samples (n = 5 HBV-HCC and n = 3 non-HBV-HCC) from the discovery cohort.



Supplementary Figure 14. Deciphering ligands-receptors interactions in HBV- and non-HBV-HCC microenvironment, related to Figure 8. A. Bar chart showing the number of significant ligand-receptor pairs among T/NK (left top panel), endothelial (left bottom panel), HSCs (right top panel), B

cells (right bottom panel), and all other cells in HBV and non-HBV tumor samples. **B.** Kaplan-Meier plots of disease-free survival in the publicly available TCGA LIHC cohort, comparing low versus high LGALS9-SORL1 ligand-receptor pair's mean expression in Hepatitis virus (bottom panel) and non-hepatitis virus (top panel) HCC patients. The number of patients from each subgroup and the Log-Rank t-test results are indicated in the plot. N represents the number of patients. **C.** Dotplot showing the interaction between T/NK clusters and malignant cells in HBV- and non-HBV-HCC, based on selected co-inhibitory ligand and receptor pairs. These scores are normalized expression levels, dot size indicates the significance of the interaction, and dot color indicates interaction intensity, calculated by CellPhoneDB. **D.** Kaplan-Meier plots of disease-free survival in the publicly available TCGA LIHC cohort, comparing low versus high CTLA4-CD86 ligand-receptor pair's mean expression in Hepatitis virus (right panel) and non-hepatitis virus (left panel) HCC patients. The number of patients from each subgroup and the Log-Rank t-test results are indicated in the plot. N represents the number of patients.