Supplementary materials

Single cell RNA sequencing analysis

Fresh neoplastic specimens were immediately processed and dissociated into single-cell suspensions. Cell capture and library preparation were conducted utilizing the Singleron GEXscope platform, with subsequent high-throughput sequencing performed on the Illumina HiSeq system. (1) Data Preprocessing and Quality Control (QC): Raw sequencing reads were processed using the CeleScope pipeline (Singleron Biotechnologies) to generate gene expression matrices. Cell quality control was performed using the SynEcoSys ™ platform according to Singleron-recommended criteria. Specifically, cells with fewer than 200 detected genes were excluded. The upper thresholds for detected genes and UMI counts were dynamically defined as the 98th percentile within each individual sample to remove potential outlier cells. Mitochondrial-content filtering was performed using an adaptive threshold selected from 5%, 10%, 15%, 20%, 30%, or 50%, with the threshold chosen to exclude approximately 5% of cells in each sample. In addition, cells with a ribosomal gene fraction $\geqslant$ 50% were removed. To minimize technical artifacts, ambient RNA contamination was corrected using DecontX (celda package, version 1.6.1) on a per-sample basis. Potential doublets were identified and removed using Scrublet (version 0.2.3). (2) Data Normalization, Dimensionality Reduction, and Clustering: Downstream analyses were performed using Scanpy (version 1.9.3) and anndata (version 0.7.8). Gene expression matrices were normalized and log-transformed using the standard SynEcoSys/Scanpy workflow. A total of 2,000

highly variable genes (HVGs) were selected for downstream analyses. Principal component analysis (PCA) was performed using a data-driven selection of principal components. The number of PCs retained satisfied the following criteria: (i) cumulative explained variance exceeded 90%, and (ii) the explained variance of an individual PC was below 5%, or the difference in explained variance between two consecutive PCs was less than 0.1%. Cell-cycle effect correction was not applied. Clustering was performed using Leiden (version 0.10.1) and Louvain (version 0.8.1) algorithms with a clustering resolution parameter of 1.2. No additional batch-effect correction algorithm was applied because no obvious sample-specific clustering pattern indicative of batch effects was observed during exploratory analyses. (3) Cell Type Annotation: Cell-type annotation was performed using an automated annotation strategy implemented in SynEcoSys™. Specifically, the CellID algorithm (version 0.1.0) was used to match cluster-specific marker genes against marker-based reference datasets within the SynEcoSys database. Human liver (Homo sapiens) reference datasets were used for annotation, and cell identities were assigned according to the best-matching marker signatures and canonical cell markers. (4) Supplementary Data Availability: The median number of cells retained per sample after quality control is provided in Supplementary Table S3.

Sample size calculation

Previous studies have reported that high fibroblast activation protein (FAP) expression is observed in approximately 30-50% of hepatocellular carcinoma (HCC)

cases. Assuming a conservative estimated proportion within this range and setting the allowable error at 10% of the estimated proportion, the calculated sample size required to achieve a 95% confidence interval was 384-576 cases. The present study included 1003 patients across the training and validation cohorts, substantially exceeding the minimum requirement and thereby ensuring sufficient statistical power for subsequent analyses.

Table S1. Histological characteristics staining

Histological staining	Antibody	Negative finding	Positive finding
FAP	Rabbit monoclonal, A23789, ABclonal Technology	Low (H-score ≤10)	High (H-score >10)
CD34	Rabbit monoclonal, IR034, LBP Medicine Science & Technology	Absence (VETC)	Presence (VETC)

Abbreviations: FAP, fibroblast activation protein; VETC, vessels encapsulating tumor clusters

Table S2. Interreader agreement of clinicopathological features

Features	Kappa
Tumor size classification	0.91 (95% CI: 0.88-0.94)
Microvascular invasion	0.85 (95% CI: 0.82-0.89)
Satellite lesions	0.93 (95% CI: 0.91-0.96)
Tumor differentiation	0.90 (95% CI: 0.87-0.93)

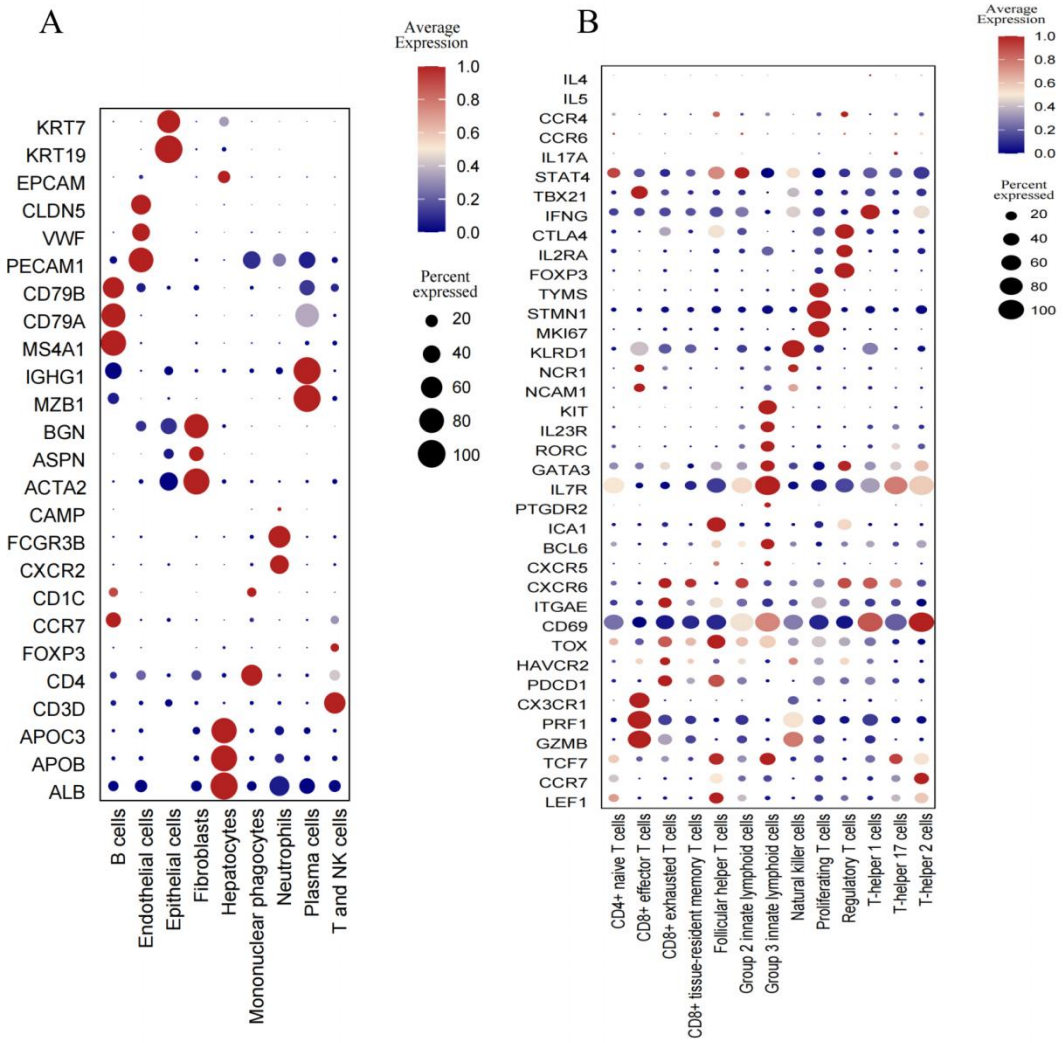
Tertiary lymphoid structure	0.92 (95% CI: 0.90-0.94)
MVD	0.93 (95% CI: 0.78-0.97)
VETC pattern	0.90 (95% CI: 0.87-0.93)
FAP expression level	0.90 (95% CI: 0.87-0.93)

Abbreviations: MVD, Microvascular density; FAP, fibroblast activation protein; VETC, vessels encapsulating tumor clusters

Table S3. Number of cells retained after quality control for each scRNA-seq sample

Cell count	Samples ID	Cell count	Samples ID
11828	CSJ1118	8172	WJQ250812003
5770	CYZ250910016	7816	WXF0328
12412	CZ251017011	7965	XMG250919011
14561	HJH1118	10270	XSZ250819005
9078	JXL251015030	4138	YSZ0401
11025	LC250828082	6512	YXG0513
7741	LHS251101008	9798	ZWS0306
8488	LHZ1202	12988	ZXH0410N
4308	LM0311	9895	ZZW250821041
8906	LXQ250828009	11803	ZZY250815025
10399	PGL251024091	6833	Zhg0327
9204	SDM250729011		

Figure S1: Transcriptional validation and canonical marker gene expression profiles for single-cell cluster annotations



(A) Dot plot illustrating the expression of classical lineage-specific marker genes used to define the 9 major cell types across the overall dataset; (B) Dot plot demonstrating the refined expression of canonical signature genes utilized to identify the 13 specific T-cell and innate lymphoid cell (ILC) subpopulations

The dot size reflects the percentage of cells expressing the target gene within a cluster, and the color intensity indicates the average relative expression level.

Figure S2: Transcriptional expression and topographic distribution of FAP in the scRNA-seq cohort.