Supplementary figure captions

Supplementary Fig. 1. Violin plots show the post-filtering quality parameters for each sample in the single-cell datasets. The yellow dashed line shows the quality control (QC) threshold. nFeatures, number of features (i.e., genes) detected in each cell. nCounts, number of counts (i.e., unique molecular identifiers) detected in each cell. \( \log_{10} \) Feature/unique molecular identifiers (UMI), \( \frac{\log_{10} \text{number of features}}{\log_{10} \text{number of UMI}} \). Mitochondrial/ribosomal/hemoglobin genes %, the percentage of all the counts belonging to mitochondrial, ribosomal or hemoglobin genes for each cell. The Y-axis does not always start at zero for the sake of presentation. (A) Quality parameters for the single-cell reference dataset of human Vδ1 and Vδ2 \( \gamma \delta \) T cells purified from adults’ peripheral blood, which were provided to the deconvolution algorithm. (B) Quality indicators for the scRNA-seq data of tumor samples from patients with cervical cancer. (C) Quality indicators for the scRNA-seq data of TIL-infused products.

Supplementary Fig. 2. The number of patients from TCGA-CESC cohort with each type of data, which were involved in the studies.

Supplementary Fig. 3. Pearson correlation between the abundance of \( \gamma \delta \) T cells and other immune cells in cervical cancerous tissue.

Supplementary Fig. 4. Overall landscape of SNVs (A-B) and CNAs (C-D) in patients with low or high abundance of \( \gamma \delta \) T cells.

Supplementary Fig. 5. The survival analysis was conducted on 287 patients in the TCGA-CESC cohort, for whom RNA-seq data, SNV or CNA data, and prognostic
information were all available. Kaplan-Meier curves depict OS (A) and PFI (B) based on the mutation status of each gene in Figure 3A. Mutations considered encompassed both SNVs and CNAs.

**Supplementary Fig. 6.** Representative pathways significantly enriched in patients with a high abundance of γδ T cells identified by GSEA based on the KEGG database.

**Supplementary Fig. 7.** Preprocessing of scRNA-seq data from cervical cancer tumor tissue samples. (A) Highly variable genes. (B) Elbow chart of principal components. (C-D) t-SNE and UMAP plots, colored by patients. (E-F) t-SNE and UMAP plots, colored by Seurat clusters.

**Supplementary Fig. 8.** (A) The top 5 highly expressed genes in each cell cluster identified by scRNA-seq of cervical cancer tissues. (B) The expression of the top 30 marker gene signatures in γδ T cells, extracted from scRNA-seq data, exhibits a significant difference between cervical cancer tissue (from the TCGA cohort, n = 306) and normal cervix tissue (from the GTEx cohort and the TCGA cohort, n = 13). (C) Dot plot showing secreted signaling communications from γδ T cells to CD8^+ αβ T cells.

**Supplementary Fig. 9.** Cell-cell communication patterns in the TME of cervical cancer. (A-B) Heatmap showing the patterns of cell-cell contacts (A) and secreted signaling communications (B) of all cell clusters in the TME of cervical cancer. (C-D) Global in-coming (C) and out-going (D) secreted signaling communication patterns of the cell cluster, determined by hierarchy clustering based on the consensus matrix.
Supplementary Fig. 10. The efficacy of immunotherapy was influenced by the abundance of γδ T cells. (A) Box plots demonstrate the expression levels of PD-L1 (left), CTLA4 (center) and IFN-γ (right) in patients with high or low infiltration levels of γδ T cells. (B) Scatter plot demonstrating a statistically significant positive correlation between the expression level of the top 30 genes signature of γδ T cells in the TME of cervical cancer and the expression level of IFN-γ. (C) Feature plots show the distribution pattern of marker genes associated with CD4+ αβ T cells (left) and CD8+ αβ T cells (right) in TILs from the infused product. (D) Frequency distribution of AUCell scores for the γδ TCR constant region in TILs from the infused product. (E) Frequency distribution of AUCell scores for CD4+ T-cell markers and CD8+ T-cell markers. (F) AUCell scores for each patient responding or not responding to the TIL treatment. Patient numbers are consistent with the original study.

Supplementary methods

Single-cell preparation, library construction and sequencing

After pretreatment of the samples, the cells were washed, counted and concentrated following the guidelines described in the Cell Preparation Guide (10x Genomics®). The prepared cell suspension was loaded into Chromium Microfluidic Chips (10x Genomics®) with 3’ chemistry. Barcoding of the cells was performed using the Chromium Controller (10x Genomics®). RNA from the barcoded cells was subsequently reverse-transcribed, and sequencing libraries were constructed using reagents from Chromium Single Cell 3’ reagent kits (10x Genomics®). Paired-end sequencing was performed with a NovaSeq 6000 (Illumina®) following the
Quality control (QC) and preprocessing of scRNA-seq data

The following QC requirements are implemented progressively for each cell in the scRNA-seq data generated in this study. 1) Number of features ≥ 200; 2) Number of read counts ≥ 500; 3) \( \frac{\text{log_{10} number of features}}{\text{log_{10} number of UMIs}} \) ≥ 0.8; 4) Mitochondrial gene percentage ≤ 15%; 5) Hemoglobin gene percentage ≤ 3%; 6) Identified by Doublet Finder (version 2.03) \(^1\) as a single cell (the doublet rate of each sample was estimated as the number of cells multiplied by \(8 \times 10^{-6}\), and homotypic doublets were allowed). Public data were subjected to similar QCs, with specific thresholds labeled on violin plots. \((\text{Supplementary Fig. 1})\) The SCTransform v2 algorithm \(^2\) was used for normalization and variance stabilization. In this process, heterogeneity associated with sequencing depth and mitochondrial contamination was regressed out. Linear dimensionality reduction was performed based on principal component analysis (PCA). Graph-based dimensionality reduction was performed using t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) algorithms. The resolution for cell clustering was determined based on Clustree \(^3\) (version 0.5). Cells were clustered using the Louvain algorithm. The dataset of γδ T cells underwent batch effect removal using Harmony \(^4\) (version 0.1.1) prior to visualization.

Annotation of cell clusters

Based on a previous study \(^5\) and databases \(^6\) \(^7\), the expression of the features including but not limited to the following, was considered the main basis for cluster
annotation. As a reference, annotations based on CellTypist (version 1.3) and singleR (version 2.0) were also performed. First, epithelial cells (expressing EPCAM and CD24), endothelial cells (expressing CDH5 and VWF), matrix fibroblasts (expressing LUM and DCN), vascular fibroblasts (expressing RGS5 and MYH11), and immune cells (expressing PTPRC) were distinguished. Among them, immune cells were further categorized into macrophages (expressing LYZ, CST3, CD68 and CD163), monocytes (expressing LYZ, CST3, FCN1 and APOBEC3A), dendritic cells (including conventional dendritic cells expressing LYZ, CST3, CD1C and HLA-DQB2, and plasmacytoid dendritic cells expressing IL3RA), neutrophils (expressing FCGR3B and CSF3R), mast cells (expressing KIT and CPA3), natural killer cells (expressing FGFBP2 and KLRF1), B cells (expressing MS4A1 and CD79B), plasma cells (expressing IGHG1 and MZB1) and T cells (expressing CD3D and CD3E). Carefully, proliferating T cells (expressing MKI67 and TOP2A) were defined separately. Next, CD4+ T cells were categorized into regulatory T cells (expressing FOXP3 and IL2RA), CD4+ effector memory T cells (expressing CD40LG) and CD4+ exhausted T cells (expressing CTLA4, PDCD1 and CD40LG). CD8+ T cells were categorized as CD8+ effector memory T cells (expressing GZMB and GZMK), CD8+ tissue-resident memory T cells (expressing CCL4L2 and KLRC1) and CD8+ exhausted T cells (expressing LAG3, PDCD1 and CTLA4). Finally, γδ T cells were distinguished (expressing TRDC, TRGC1 and TRGC2 highly, expressing CD8A and CD8B very lowly, and not expressing CD4, FGFBP2 or KLRC1). Cellular clusters were further annotated based on the top expressed marker genes.
Detailed parameters for the deconvolution

Quantile normalization was disabled, and a minimum average gene expression threshold of $\log_2{0.25}$ was implemented. The reference was constructed by performing 5 replicates and selecting 50% of the available single cell gene expression profiles through random sampling. The kappa value for the signature matrix was set to 999.

For the differential expression analysis, a q-value cutoff of 0.01 was used. The lower and upper thresholds for the number of barcode genes were defined as 300 and 500, respectively. Batch correction was conducted using S-mode, and a total of 100 permutations were performed for subsequent statistical analysis.

Differential expression analyses

Count matrices obtained from STAR\textsuperscript{10} were used for bulk transcriptome analysis, while transcripts per million (TPM) normalized matrices were utilized in downstream analyses as applicable. Gene differential expression analyses of bulk expression profiling were performed using the limma\textsuperscript{11} package (version 3.56.2), with significance thresholds set to a p-value less than 0.05 and a $\log_2$FC larger than 1. Differentially expressed genes between two groups of cells were identified using a Wilcoxon rank-sum test, and the hyperbolic method was used to delimit the significance threshold.\textsuperscript{12} In addition, the expression of a multi-gene signature was analyzed using GEPIA2\textsuperscript{13}.

Gene set analyses

For the scRNA-seq dataset of in vitro-expanded TILs, the scores of cell-specific gene sets (GS) were computed using AUCell.\textsuperscript{14} Three GS were involved in the
scoring. GS\textsubscript{1}: TRDC, TRGC\textsubscript{1}, TRGC\textsubscript{2} (for γδ T cells); GS\textsubscript{2}: CD4, CD40LG (for CD4\textsuperscript{+} T cells); GS\textsubscript{3}: CD8A, CD8B (for CD8\textsuperscript{+} T cells). Gene set level analyses were conducted based on the gene set variation analysis (GSVA)\textsuperscript{15} or gene set enrichment analysis (GSEA)\textsuperscript{16} algorithms. The clusterProfiler\textsuperscript{17} package (version 4.8.2) was used for the implementation and visualization of the analyses. For GSVA analyses, gene sets from the Gene Ontology (GO)\textsuperscript{18,19} database were employed. Differential analysis of gene set scores was performed using the limma\textsuperscript{11} package (version 3.56.2). For GSEA analyses, the hallmark gene set from the Molecular Signatures Database (MSigDB, version 2023.1.Hs updated March 2023)\textsuperscript{20} and biological pathway data from Kyoto Encyclopedia of Genes and Genomes (KEGG)\textsuperscript{21} were utilized.

**Preprocessing of MRI images and radiomic feature extraction**

Patients were divided into a training set including 40 patients and a test set including 13 patients. Gross tumor volume was manually segmented as the volume of interest (VOI) by an experienced radiation oncologist and subsequently reviewed by a senior physician. A comprehensive collection of 824 features was compiled, incorporating 14 shape-based, 17 intensity histogram-based, and 73 texture-based features. To further enhance the feature set, wavelet decomposition was conducted at 8 levels, and these features were recalculated for each possible combination of a high or low pass filter applied to each of the three cardinal axes, resulting in 720 additional features. Before the feature extraction process, isotropic spatial resampling was performed on the MRI images, setting a resolution of 1 mm. To address the challenge of working with MRI images that have arbitrary intensity units, intensity outlier
filtering was applied. Specifically, gray values within the VOI that fell outside the range of mean ± 3 standard deviations were excluded from the analysis. Additionally, discretization was performed using a fixed bin number of 100.

Appendix

The formula for measuring γδ T-cell abundance based on the radiomics model

Predicted γδ T cell abundance

\[
= 0.2371 + 0.3706 \times \text{original\_shape\_Flatness} + 0.1549 \\
\times \text{wavelet.LLH\_firstorder\_Median} + 0.1937 \\
\times \text{wavelet.LHH\_glcm\_ClusterShade} - 0.6241 \\
\times \text{wavelet.HLL\_firstorder\_Median} + 0.9586 \\
\times \text{wavelet.HLL\_firstorder\_RootMeanSquared} + 0.2817 \\
\times \text{wavelet.HLL\_glcm\_Correlation} + 0.311 \\
\times \text{wavelet.HHL\_firstorder\_Mean} + 0.1211 \\
\times \text{wavelet.HHL\_firstorder\_Median}
\]

Supplementary references


17. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021;2(3):100141. doi:


