RESEARCH ARTICLE

scSurvival: Single-Cell Survival Analysis of Clinical Cancer Cohort Data at Cellular Resolution

Tao Ren[1,2], Faming Zhao[1,2], Canping Chen[1,2], Le Zhou[1,2], Ling-Yun Wu[3,4], Gordon B. Mills[2,5], Lisa M. Coussens[2,6], and Zheng Xia[1,2,7]

ABSTRACT Survival analysis is fundamental to cancer research. Advances in technology have enabled an increasing number of cohort-level cancer studies to incorporate single-cell sequencing alongside clinical survival data. However, no effective strategy currently exists for directly modeling survival outcomes from single-cell data. To address this gap, we present scSurvival, an attention-based multiple-instance Cox regression framework that models each tumor sample as an ensemble of cells to predict survival outcomes at both the patient and single-cell levels. To handle high dimensionality, sparsity, and batch effects, scSurvival integrates a variational autoencoder–based feature extraction module with generative modeling to enhance feature robustness and cross-batch generalizability. Comprehensive simulations demonstrate scSurvival's superior performance and scalability. In melanoma and liver cancer single-cell RNA sequencing (scRNA-seq) cohorts, scSurvival accurately predicts patient outcomes and identifies the cell subpopulations most critical to survival. Overall, scSurvival enables robust prediction of patient survival while uncovering survival-associated cell subpopulations, advancing single-cell survival analysis in cancer research.

SIGNIFICANCE: Survival analysis is central to clinical oncology, yet no effective tools currently model survival outcomes directly from single-cell data. scSurvival bridges this gap by predicting patient outcomes and identifying key subpopulations from scRNA-seq with survival information, enabling scalable analyses and promoting broader adoption of cohort-level single-cell profiling in cancer research.

INTRODUCTION

Survival analysis has long been a cornerstone of cancer research for investigating time-to-event outcomes such as overall survival (OS) and progression-free survival (PFS). Bulk RNA sequencing (RNA-seq) studies have demonstrated the versatility of survival analysis by modeling gene expression as covariates, constructing multigene risk scores, and stratifying patients into molecular subgroups for prognostic evaluation and assessment of therapeutic efficacy (1–4). Meanwhile, single-cell RNA-seq (scRNA-seq) provides unprecedented resolution for characterizing the cellular heterogeneity of the tumor microenvironment (TME; refs. 5–8). Leveraging this capability, recent advancements and the widespread adoption of scRNA-seq have led to a growing number of cohort-level cancer studies that profile single-cell transcriptomes from hundreds of patients alongside clinical survival data, offering new opportunities to elucidate how cellular heterogeneity influences disease progression and patient outcomes. Because individual cells can influence patient outcomes differently, identifying survival-associated subpopulations is crucial for improving risk assessment and developing more precise and targeted therapies beyond bulk-level analyses.

In early studies, because of the limited sample size of scRNA-seq data in cancer studies, the predominant strategy for identifying risk-associated cell populations has involved transferring survival information from existing bulk RNA-seq datasets to single-cell data by integrating single-cell and bulk data, such as Scissor (9), scSurv (10), and DEGAS (11). Although these approaches have enabled the identification of cell populations associated with survival outcomes, they all rely on bulk transcriptomics and its associated survival outcomes as a bridge without employing the true survival information of the single-cell data.

As cohort-level cancer studies increasingly generate scRNA-seq datasets from hundreds of patients with matched survival information, two simple strategies can be adopted to adapt traditional Cox regression frameworks to single-cell data: (i) aggregating cell-level gene expression into pseudobulk profiles for each patient and (ii) annotating major cell types and using their proportions as patient-level covariates. Although both approaches allow the use of standard survival models, they suffer from major information loss. The pseudobulk strategy treats all cells equally and obscures the contributions of rare but prognostically important subpopulations. In contrast, the cell type proportion strategy assumes homogeneity within annotated cell types and ignores intratype functional diversity. These methods offer a coarse approximation but fall short of realizing the full potential of single-cell data in capturing fine-grained risk structures. To our knowledge, directly performing survival prediction at single-cell resolution,

[1]Biomedical Engineering Department, Oregon Health & Science University, Portland, Oregon. [2]Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon. [3]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. [4]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, China. [5]Division of Oncological Sciences, Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon. [6]Cell, Development and Cancer Biology Department, Oregon Health & Science University, Portland, Oregon. [7]Center for Biomedical Data Science, Oregon Health & Science University, Portland, Oregon.

T. Ren and F. Zhao contributed equally to this article.

Corresponding Author: Zheng Xia, Knight Cancer Institute, Oregon Health & Science University, 3181 SW Sam Jackson Park Rd., Portland, OR 97239. E-mail: xiaz@ohsu.edu

Cancer Discov 2026;16:931–52

doi: 10.1158/2159-8290.CD-25-0965

while preserving intrapatient heterogeneity and relying solely on single-cell cohort data, remains an unmet challenge, and no dedicated computational tools have yet been developed to address this need.

Constructing risk prediction models such as the Cox proportional hazards model directly from single-cell cohort data also fundamentally differs from conventional bulk data–based Cox regression. In traditional bulk settings, each patient's covariate values form a single feature vector that can be directly used for model construction. In contrast, single-cell expression data are structured as a cell-by-gene matrix rather than a simple per-patient vector, with individual cells contributing differently to patient outcomes. Therefore, there is a clear need for a framework that can simultaneously predict patient survival outcomes and identify the most informative cell subpopulations. Attention-based multiple-instance learning (12–14) offers a compelling solution for this task. In this framework, each tumor sample is treated as a collection of multiple instances (i.e., cells). The model learns to predict patient-level labels while simultaneously assigning weights to individual instances, thereby highlighting the cell subpopulations most critical to the survival outcomes.

In this study, we conceptualize each tumor sample as an ensemble of cellular instances and thus develop a new tool, scSurvival, specifically designed to address heterogeneous cell subpopulations in survival analysis of single-cell cancer cohort data. To tackle the high dimensionality, sparsity, noise, and severe batch effects inherent in single-cell datasets, we design a feature extraction module based on variational autoencoders (VAE; arXiv 1312.6114) and generative modeling (15, 16). Coupled with attention-based multiple-instance Cox regression (AMICR), scSurvival enables survival modeling of patient cohorts spanning multiple experimental batches. scSurvival not only integrates single-cell expression data with optional clinical covariates to construct accurate risk prediction models but also identifies outcome-associated cell subpopulations and characterizes their risk tendencies, thereby facilitating more refined downstream biological and clinical analyses.

RESULTS

Overview of scSurvival

We developed scSurvival, a single-cell survival outcome prediction tool based on attention-based multiple-instance learning, which accounts for the varying contributions of individual cells to survival outcomes. It is designed to construct survival prediction models from single-cell cancer cohort data while simultaneously identifying cell subpopulations that are strongly associated with patient risk. The core principle involves compressing each patient's high-dimensional single-cell data into low-dimensional features, aggregating these features into patient-level representations through attention weights, and subsequently building a survival model at the patient level using Cox regression analysis. The attention mechanism preserves cellular heterogeneity within each patient, allowing subpopulations with higher attention scores to be more closely linked to survival probability. By aligning cell-level features with aggregated patient-level representations, the survival prediction model can directly infer the risk tendencies of specific cell subpopulations, thereby enabling the identification of risk-associated cells.

Specifically, the main inputs to scSurvival include single-cell expression matrices measured for each patient (Fig. 1A and B) and patient survival information, i.e., survival event status and event time (Fig. 1C). Conceptually, scSurvival comprises two key modules: a cell feature extraction module (Fig. 1D) and an AMICR module (Fig. 1E). The feature extraction module is a VAE-based generative model that captures technical dropout using a zero-inflated Gaussian (ZIG) distribution (17, 18) and models the log-normalized expression matrix. LayerNorm [LN; (19)] and squeeze-and-excitation (SE; ref. 20) modules are integrated into this architecture to enable adaptive standardization and reweighting of gene features. In addition, the feature extraction module supports batch label inputs to extract cross-batch single-cell features (Fig. 1D). The AMICR module aggregates cell-level features into patient-level representations using a multi-head attention mechanism and inputs the aggregated features into a risk scorer to perform hazard scoring and Cox regression analysis (Fig. 1E), in which the structures of the cell-level attention mechanism, gene-reweighting SE module, and risk scorer are shown in Fig. 1F. Furthermore, patient-level clinical covariates (21) can be incorporated into the risk scorer for joint analysis (Fig. 1G), enabling appropriate control of confounding influences and thereby improving the accuracy of survival prediction models and the identification of risk-associated cell subpopulations.

The final loss function of the model comprises three components: the negative log-likelihood of Cox regression, the reconstruction likelihood of the ZIG-VAE, and an entropy regularization term (22) to control the sparsity of the cell-level attention weights. Training is performed in two stages: pretraining and fine-tuning (Fig. 1H). In the pretraining stage, only the VAE module is trained to extract stable cell features. During the fine-tuning stage, all three loss components are jointly optimized to facilitate multiple-instance Cox regression, producing optimized cell representations and attention weights that are adaptively aligned with survival outcomes ("Methods").

The resulting outputs of scSurvival are the attention-adjusted hazard score for each cell (*hazard_adj*; see "Methods", Fig. 1I), along with patient-level risk scores. This represents a fundamental distinction from traditional survival analysis methods: scSurvival not only provides risk stratification at the patient level but also infers intrapatient cellular risk heterogeneity. Given the hazard scores of individual cells estimated by scSurvival, we can then perform downstream analyses of heterogeneous cell states within each cell type using other tools, such as identifying signature genes and altered pathway activities by comparing higher- and lower-hazard subpopulations (Fig. 1J). Furthermore, scSurvival functions as a survival prediction model based on single-cell expression data, enabling risk assessment for new patients (Fig. 1K).

Benchmarking scSurvival for Hazard Score Prediction at Single-Cell Resolution

To systematically evaluate the effectiveness of scSurvival, we generated a series of simulated datasets and assessed its performance in controlled experiments. We first used Splatter

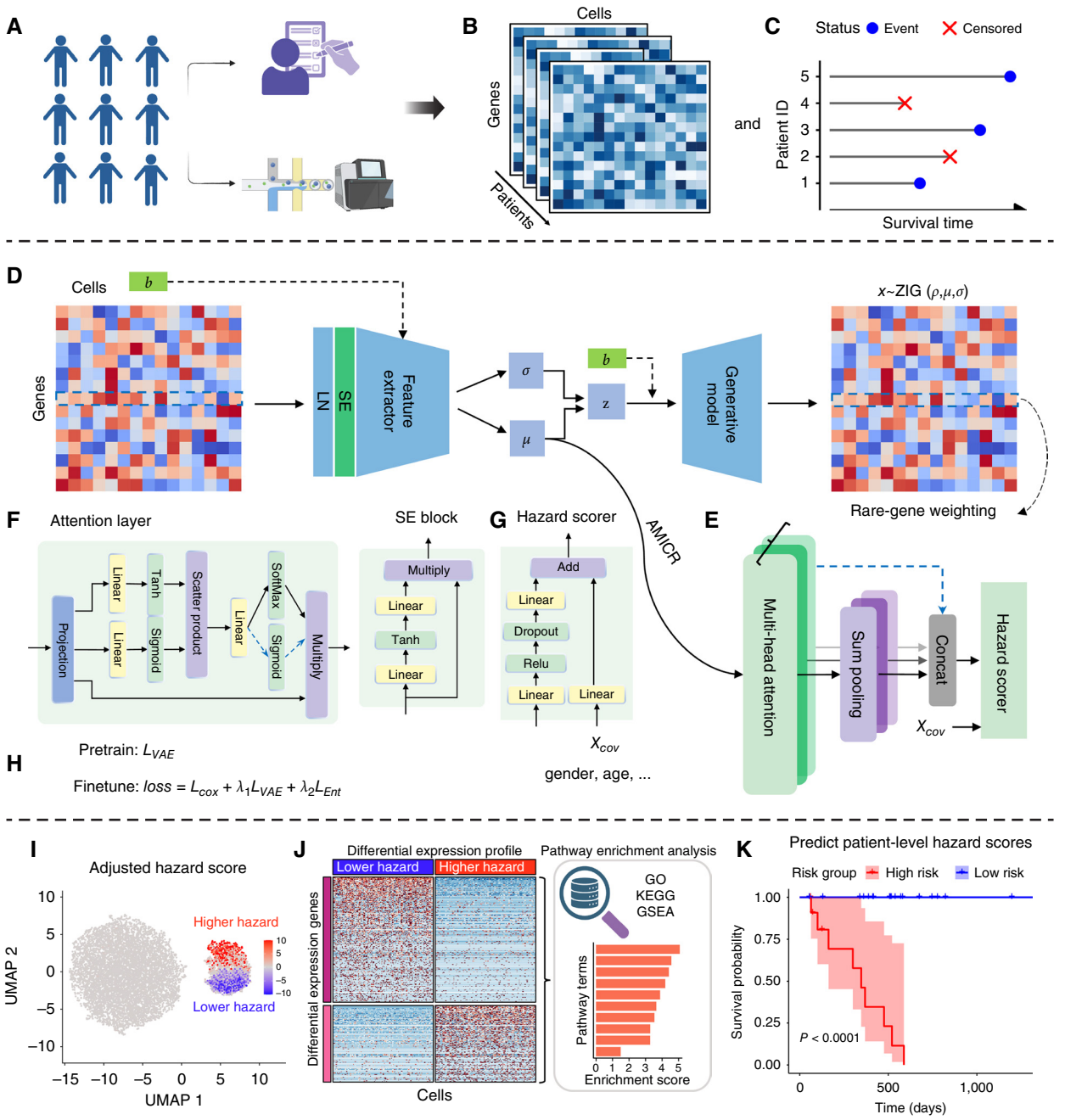

Figure 1. Framework of scSurvival. **A,** Generation of single-cell cohort data: each patient undergoes scRNA-seq profiling, with corresponding clinical follow-up records. **B,** Single-cell transcriptomic expression matrix for each patient. **C,** Survival information for each patient, including event status and event time. **D–H,** Model architecture of scSurvival. **D,** A single-cell feature extraction module based on a ZIG-VAE, which takes log-normalized expression matrices as input and extracts batch-integrated single-cell features. LN and SE blocks are incorporated to adaptively normalize gene feature scales. **E,** A multi-head AMICR module that aggregates single-cell features to the patient level and computes relative hazard scores. **F,** The detailed architectures of the attention layer, the SE block in the VAE encoder, and the hazard scorer. **G,** The hazard scorer supports the optional inclusion of additional covariates (e.g., gender and age) if available. **H,** The overall loss function of scSurvival and the two-stage training strategy: the first stage involves pretraining only the VAE module to extract stable features and the second stage jointly optimizes the entire model to perform multiple-instance Cox regression and risk cell selection. **I,** Cell hazard scores estimated by scSurvival after adjustment using attention weights. **J,** Differential expression and pathway enrichment analyses of higher-hazard vs. lower-hazard cells identified by scSurvival. **K,** scSurvival can predict hazard scores for new patients in independent datasets to stratify patients into high- and low-risk groups.

(23) to simulate a single-cell dataset comprising three groups (good.survival, bad.survival, and other) as the ground truth (Fig. 2A). Using these cells as risk-driving populations, we sampled single-cell expression data for 100 patients by varying the proportions of good.survival and bad.survival cells and simulated the corresponding survival times based on these

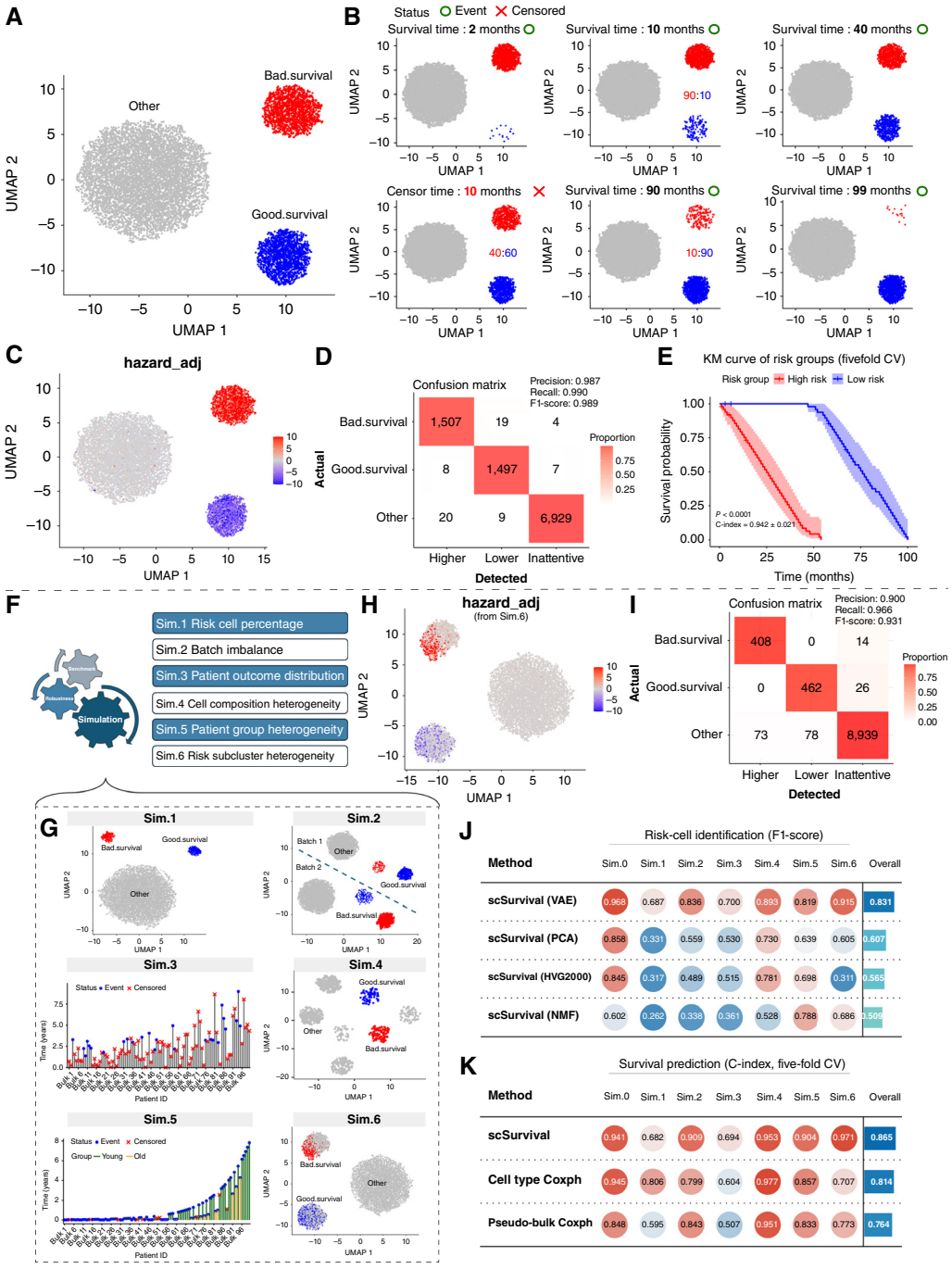

Figure 2. Performance evaluation of scSurvival on simulated datasets. **A,** UMAP showing simulated scRNA-seq data with survival-driving cell subpopulations as the ground truth, including three groups: good.survival, bad.survival, and other background cells. **B,** UMAP plots of simulated scRNA-seq data from patients with varying survival times and differing proportions of cell subpopulations. Patients with longer survival times had higher proportions of good.survival cells and lower proportions of bad.survival cells. **C,** UMAP showing cell hazard scores estimated by scSurvival. **D,** Confusion matrix and performance evaluation for ground-truth cell subpopulation identification based on adjusted hazard scores predicted by scSurvival. **E,** Fivefold CV for patient-level survival prediction: merged relative risk percentiles across folds were used to evaluate risk stratification, with significance assessed by log-rank tests; the mean C-index and its SD across folds were also reported. **F,** Overview of six simulation scenarios designed to evaluate the performance and robustness of scSurvival under different conditions. **G,** UMAP or survival annotation illustrations for each simulation setting (Sim.1–Sim.6), highlighting the distinct designs and features used to introduce biological or technical variability. **H,** UMAP showing cell-level hazard scores (hazard_adj) estimated by scSurvival in Sim.6 (risk subcluster heterogeneity), in which true risk cells only form a small subcluster within a larger cluster. **I,** Confusion matrix showing the identification performance of scSurvival in Sim.6. **J,** Comparison of F1-score (mean of 10 repeats) for ground-truth risk cell identification across different simulation scenarios and feature extraction strategies, including scSurvival with the default VAE and alternative representations based on PCA, HVG2000, and NMF. Sim.0 refers to the baseline simulation setting shown in **A** and **B**, used as the basic scenario without added complexity. **K,** Comparison of patient-level survival prediction performance (mean C-index from five-fold CV) between scSurvival and two baseline models: Cell type Coxph and pseudobulk Coxph across all simulation settings. Sim.0 refers to the baseline setting illustrated in **A** and **B**. KM, Kaplan–Meier.

proportions, with random censoring applied to a subset of patients to mimic event truncation (Fig. 2B; "Methods"). We referred to this basic simulation as Sim.0. The cell attention scores and hazard scores inferred by scSurvival (Supplementary Fig. S1A; Fig. 2C) accurately delineated the ground-truth risk-driving cell subpopulations. In one experiment, cell grouping based on attention and hazard scores (Supplementary Fig. S1B; "Methods") identified 99% of the ground-truth driver cells, with an overall accuracy of 98.7% and an F1-score of 0.989 (Fig. 2D). With regard to patient-level risk prediction, fivefold cross-validation (CV) showed that the hazard scores predicted by scSurvival significantly stratified patient survival outcomes (log-rank test, P value < 0.0001), achieving a mean concordance index (C-index) of 0.942 (SD, 0.021) across folds (Fig. 2E; Supplementary Fig. S1C).

To assess whether scSurvival can accurately handle batch effects, we used Splatter to simulate ground-truth single-cell datasets from two distinct batches (Supplementary Fig. S1D; "Methods"). scSurvival successfully produced cell-level risk assessments consistent with the ground truth (Supplementary Fig. S1E), demonstrating the capability of its embedded feature extraction module to correct batch effects (Fig. 1D). Uniform Manifold Approximation and Projection (UMAP) visualization based on the latent embeddings learned by scSurvival confirmed effective batch effect removal and accurate estimation of risk tendencies (Supplementary Fig. S1F and S1G). In this batch effect simulation example, scSurvival achieved an accuracy of 97.9%, a recall of 98.2%, and an F1-score of 0.980 for ground-truth risk-driving cell identification (Supplementary Fig. S1H). Repeated evaluations across 20 runs further confirmed the model's robustness in both batch-free and batch-affected settings, with high average F1-scores (0.964 vs. 0.923; Supplementary Fig. S1I).

The performance of a survival prediction model is always affected by the sample size (cohort size) and the strength of association between covariates and survival time (risk effect size). To evaluate this, we simulated datasets with varying cohort sizes and risk effect sizes ("Methods"). The results showed that when the cohort size reached 10 patients and the effect size was moderate or stronger, the model consistently identified the risk-driving cell subpopulations, with F1-scores significantly higher than those observed under weaker effect sizes or smaller sample sizes (Supplementary Fig. S1J). In addition, when the cohort size was small, increasing the number of cells per patient also improved identification performance (Supplementary Fig. S1K).

Building upon Sim.0, we further designed six more complex simulation scenarios (Sim.1–Sim.6, Fig. 2F), systematically introducing various biological and technical confounding factors at different levels of perturbation to evaluate the robustness and performance of scSurvival across diverse data conditions (Fig. 2F and G; "Methods"). The six scenarios included variation in the percentage of risk cells (Sim.1), interbatch imbalance in risk cell distribution (Sim.2), differences in patient outcome distributions (varying effect size and censoring rate, Sim.3), cell composition heterogeneity (Sim.4), patient group heterogeneity in which subgroups respond differently to risk effects (Sim.5), and subcluster-level risk in which the risk-driving cells are embedded within larger cell clusters (Sim.6). Across most settings, scSurvival consistently identified the true risk-driving cell subpopulations, with average F1-scores typically above 0.8 (Supplementary Fig. S2A–S2F). In certain challenging scenarios, such as when the risk cell proportion was as low as 1% (Sim.1), performance declined but rapidly recovered as signal strength increased, demonstrating strong trend stability and generalizability. Notably, in the most challenging setting (Sim.6), in which the true risk cells were nested within a large cluster and lacked clear boundaries, scSurvival still accurately recovered the target population (Supplementary Fig. S2F; Fig. 2H), achieving a precision of 90%, recall of 96.6%, and F1-score of 0.931 in a representative example (Fig. 2I), highlighting its ability to perform fine-grained analysis without relying on predefined clustering. In patient-level survival prediction tasks, scSurvival also maintained consistent and stable performance across all simulation conditions (Supplementary Fig. S2G–S2L), with C-index values remaining high across fivefold CV (mostly >0.8).

Finally, we constructed benchmark comparisons on representative examples from each simulation setting ("Methods") to systematically evaluate both the feature extraction module and the overall algorithmic performance of scSurvival. For feature representation, we compared our VAE-based joint learning framework with three alternative strategies: principal component analysis (PCA), highly variable gene selection (HVG2000), and nonnegative matrix factorization (NMF). Joint learning of VAE consistently achieved the highest F1-scores across all simulation scenarios (Fig. 2J), highlighting its advantage in capturing complex expression structures and risk signals. At the algorithmic level, we further compared scSurvival with two commonly used baseline approaches: Cox regression based on pseudo-bulk expression profiles (pseudobulk Coxph) and Cox regression based on cell type proportions (cell type Coxph). Because both baseline methods lack cell-level resolution, comparisons were restricted to patient-level survival prediction. Across most scenarios, scSurvival outperformed the baselines, achieving higher C-index values in fivefold CV (Fig. 2K). Together, these comprehensive simulation results demonstrate that scSurvival offers superior capabilities in both risk subpopulation identification and survival prediction, while fully preserving single-cell resolution.

scSurvival Is Fast and Scalable

Cohort-level cancer studies often involve sequencing large numbers of patients, generating datasets with millions of cells. Performing survival analysis at this scale poses significant computational challenges for algorithm design. Benefiting from the supervised learning architecture of scSurvival, we can fully leverage the computational power of high-performance graphics processing units to process datasets of this magnitude. To evaluate its computational performance, we simulated a series of datasets and measured runtime, memory usage, and GPU memory consumption. In these tests, early stopping was disabled, and the fine-tuning phase of each run was set to complete all 500 maximum epochs ("Methods"), suggesting that in practical applications, scSurvival could handle similarly sized datasets in even shorter times.

We simulated two scenarios to evaluate the scalability of scSurvival, keeping the number of genes fixed at 2,000 in all cases, consistent with the top 2,000 most variable gene

(MVG2000) feature selection strategy commonly used in real analyses. First, we fixed the cohort size at 100 patients and gradually increased the total number of cells from 5,000 to 1 million, distributing cells evenly across patients (Fig. 3A). The runtime and memory usage of scSurvival exhibited linear growth relative to the number of cells (*x*-axis in log scale), completing the analysis of 1 million cells in approximately 17.5 minutes, with around 30 GB of memory and 50 GB of GPU memory usage. We also compared cases with and without batch effect adjustment, treating each patient as a distinct batch. The results showed that batch correction did not introduce significant additional computational burden. Next, we fixed the total number of cells at 100,000 and simulated cohort sizes ranging from 10 to 500 patients (Fig. 3B). Runtime again exhibited linear scaling with respect to the number of patients (*x*-axis in log scale), whereas memory usage remained nearly constant. GPU memory usage increased slightly when batch correction was included but remained stable without it. In the largest simulated cohort of 500 patients, scSurvival completed the analysis in approximately 33 minutes, with a peak memory usage of about 3 GB and a peak GPU memory usage of about 6 GB.

All tests were performed on an Arm64 CPU and H100 GPU with 96 GB VRAM. As different hardware configurations can lead to substantial performance variations, we also benchmarked scSurvival on the A100 platform. The computational cost remained within an acceptable range, approximately 2.5 hours for 1 million cells across 100 patients and about 1.85 hours for 100,000 cells across 500 patients (Supplementary Fig. S3A and S3B). These results demonstrate the scalability of scSurvival for processing ultra-large single-cell datasets.

scSurvival Identifies Risk-Associated Cell Subpopulations in a Melanoma Cohort

For real-world evaluation, we first applied scSurvival to a melanoma cohort dataset (24) treated with immunotherapy, comprising 48 samples from 32 patients (Supplementary Fig. S4A; Supplementary Table S1) with a total of 16,291 immune cells annotated according to the original study (Fig. 4A; ref. 24). The key cell populations, together with their attention scores and hazard scores inferred by scSurvival, are shown in Fig. 4B and Supplementary Fig. S4B and S4C. Among them, B cells and plasma cells generally exhibited lower hazard scores (Fig. 4C), a finding potentially linked to the formation of tertiary lymphoid structures (TLS; refs. 26, 27). This association was further supported by spatial transcriptomics validation, in which lower-hazard regions identified by the trained scSurvival model corresponded to TLS locations on two spatial slices (Supplementary Fig. S5A and S5B; ref. 28). By contrast, monocytes and macrophages were largely retained and showed higher overall hazard scores, but after further subgrouping of these cells ("Methods"), a bimodal pattern of hazard scores was observed, characterized by a clear polarization into higher-hazard and lower-hazard cell subsets (Fig. 4C). This observation was of particular interest as monocytes/macrophages are known to polarize into two major functional states—M1-like (antitumor) and M2-like (protumor) phenotypes (29). However, recent single-cell studies suggest that M1 and M2 phenotypes define the possible extremes of polarized cells *in vitro* but may not capture the full spectrum of tumor-associated macrophages (TAM) *in vivo* (25). We therefore sought to evaluate whether scSurvival could reliably distinguish protumor and antitumor monocyte/macrophage populations by stratifying them into higher- and lower-hazard cell subsets.

Reclustering analysis of the monocyte/macrophage compartment revealed that higher- and lower-hazard cells tend to cluster separately (Fig. 4D), suggesting transcriptional and potentially lineage similarity within each risk group. Interestingly, we also observed that some subclusters, such as cluster 1, contained a mixture of both higher- and lower-hazard cells (Fig. 4D), indicating intracluster heterogeneity in risk states. Notably, the distribution of M1 and M2 gene signature scores showed strong concordance with our predicted risk states, further supporting the biological relevance of the subpopulations identified by scSurvival (Fig. 4D; Supplementary Fig. S5C). To further explore transcriptional differences between these cell populations, we performed differential expression analysis comparing higher-hazard and lower-hazard cells. As a result, 169 genes were significantly upregulated and 397 genes were significantly downregulated in higher-hazard cells relative to lower-hazard cells (Fig. 4E; Supplementary Table S2). Notably, several key M2-associated markers (30), including *SPP1* and *MSR1*, were among the upregulated genes in higher-hazard cells. Conversely, several canonical M1-associated markers (e.g., *IDO1* and *CXCL9*) and MHC class II genes (e.g., *HLA-DOB*) were significantly downregulated in these cells. Of particular interest, *SPP1* and *CXCL9* ranked among the top upregulated and downregulated differentially expressed genes (DEG), respectively (Fig. 4E; Supplementary Table S2). A recent study (25) has identified *CXCL9* and *SPP1* as key markers of macrophage polarization. Specifically, the *CXCL9*:*SPP1* ratio was shown to reflect antitumor immune abundance and responsiveness to immunotherapy (25). Consistent with this, we observed that lower-hazard monocytes/macrophages predominantly exhibited an $SPP1^{-}$ $CXCL9^{+}$ phenotype, whereas higher-hazard cells were characterized by an $SPP1^{+}$ $CXCL9^{-}$ profile (Fig. 4F; Supplementary Fig. S5D). Furthermore, Gene Ontology (GO) enrichment analysis revealed that DEGs upregulated in higher-hazard cells were enriched in pathways associated with neutrophil migration and chemotaxis and several lipid metabolism signals such as lipid localization and transport (Fig. 4G). In contrast, DEGs upregulated in lower-hazard cells were enriched in pathways related to immune activation, including lymphocyte differentiation and activation, type II interferon response, and MHC class II antigen presentation (Fig. 4G). These enriched pathways are consistent with known features of TAMs, in which lipid metabolism and neutrophil-recruiting chemokines are commonly linked to protumor phenotypes, whereas antigen presentation and interferon responses are hallmarks of antitumor macrophage activation (bioRxiv 2024.12.18.629070; refs. 30–32). Importantly, these biological distinctions between protumor and antitumor myeloid states were also recapitulated in an independent melanoma dataset treated with immunotherapy [GSE221553; ref. (33); Supplementary Fig. S6A–S6H], further supporting the robustness of scSurvival in integrating survival information to distinguish functionally distinct monocyte/macrophage populations.

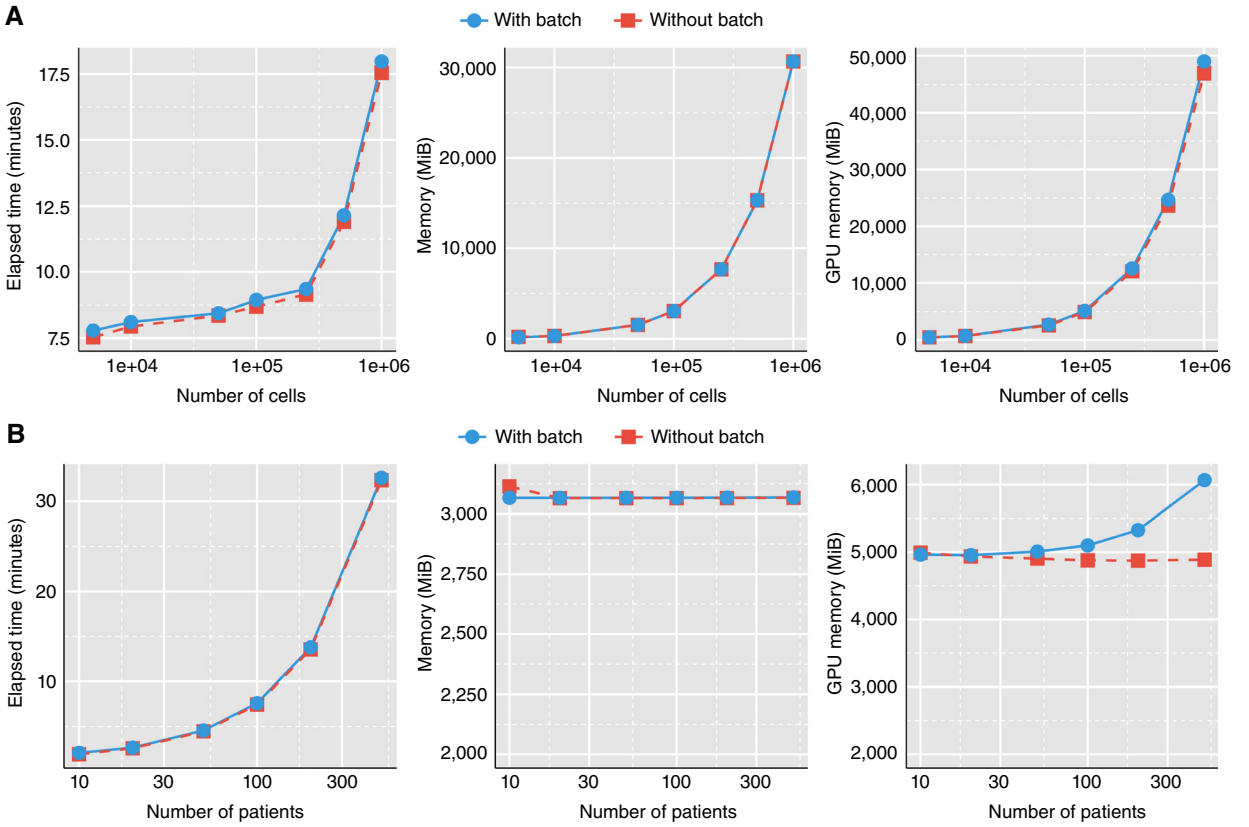

Figure 3. Runtime and resource usage of scSurvival on the H100 high-performance GPU computing platform. **A,** Runtime and memory usage of scSurvival for simulated cohorts of 100 patients across varying total numbers of cells, with or without simulated batch effects. **B,** Runtime and memory usage of scSurvival for simulated cohorts with a fixed total of 100,000 cells across varying numbers of patients, with or without simulated batch effects.

Besides, based on morphologic assessments of immunotherapy response, the original study classified the 48 samples from 32 patients into 17 responders and 31 nonresponders (24). As scSurvival enables the assessment of cellular risk heterogeneity within each patient, it revealed intrapatient variability in risk states. Overall, the key cell subpopulations defined by scSurvival (excluding inattentive cells) and their associated hazard scores showed strong concordance with immunotherapy response: key cells in responder samples had significantly lower hazard scores than those in nonresponder samples (Fig. 4H; Supplementary Fig. S7A and S7B). The proportions of lower- and higher-hazard cells within each sample further demonstrated that responder samples contained significantly more lower-hazard cells, consistent with the findings of the original study (Fig. 4I; ref. 24). Figure 4J provides an intuitive visualization of this refined cellular risk heterogeneity. Notably, among the four patients who had both responder and nonresponder samples, three patients (P28, P5, and P4) exhibited cellular risk assessments consistent with their response status (Fig. 4J; Supplementary Fig. S7C). By preserving risk heterogeneity at the cellular level, scSurvival also enables the reconstruction of sample-level summaries that integrate immune composition, survival information, and model attention. Supplementary Figure S7D summarizes these multidimensional outputs across all samples, providing an in-depth overview that highlights key immune features underlying patient-level risk.

At the patient level, beyond assessing immunotherapy response, we also conducted leave-one-out CV to evaluate the predictive performance of the scSurvival-derived risk model based on single-cell expression data. The results demonstrated that the predicted patient hazard scores significantly stratified patient survival outcomes ($P = 0.033$) and accurately ranked risks (C-index = 0.812; Fig. 4K), indicating that scSurvival established a reliable survival prediction model at the patient level. Furthermore, univariate Cox regression confirmed that scSurvival-derived hazard scores were significantly associated with OS ($P = 0.022$), with a hazard ratio comparable with cytotoxicity and more significant than other conventional immune features, including the average expression ratio of *CXCL9* and SPP1 (Supplementary Fig. S8A). The achieved C-index of 0.812 also outperformed benchmark models based on cell type fractions, pseudobulk expression, or the *CXCL9*/*SPP1* expression ratio (Supplementary Fig. S8B).

scSurvival Identifies Prognostic T-Cell States and Enables Survival Prediction in Melanoma

T cells play a crucial role in tumor immunotherapy. To better highlight the role of T cells, we isolated T cells (10,685 cells) from the previous melanoma dataset (Fig. 5A) and analyzed them with scSurvival. To more rigorously assess the predictive power of scSurvival, we further evaluated a

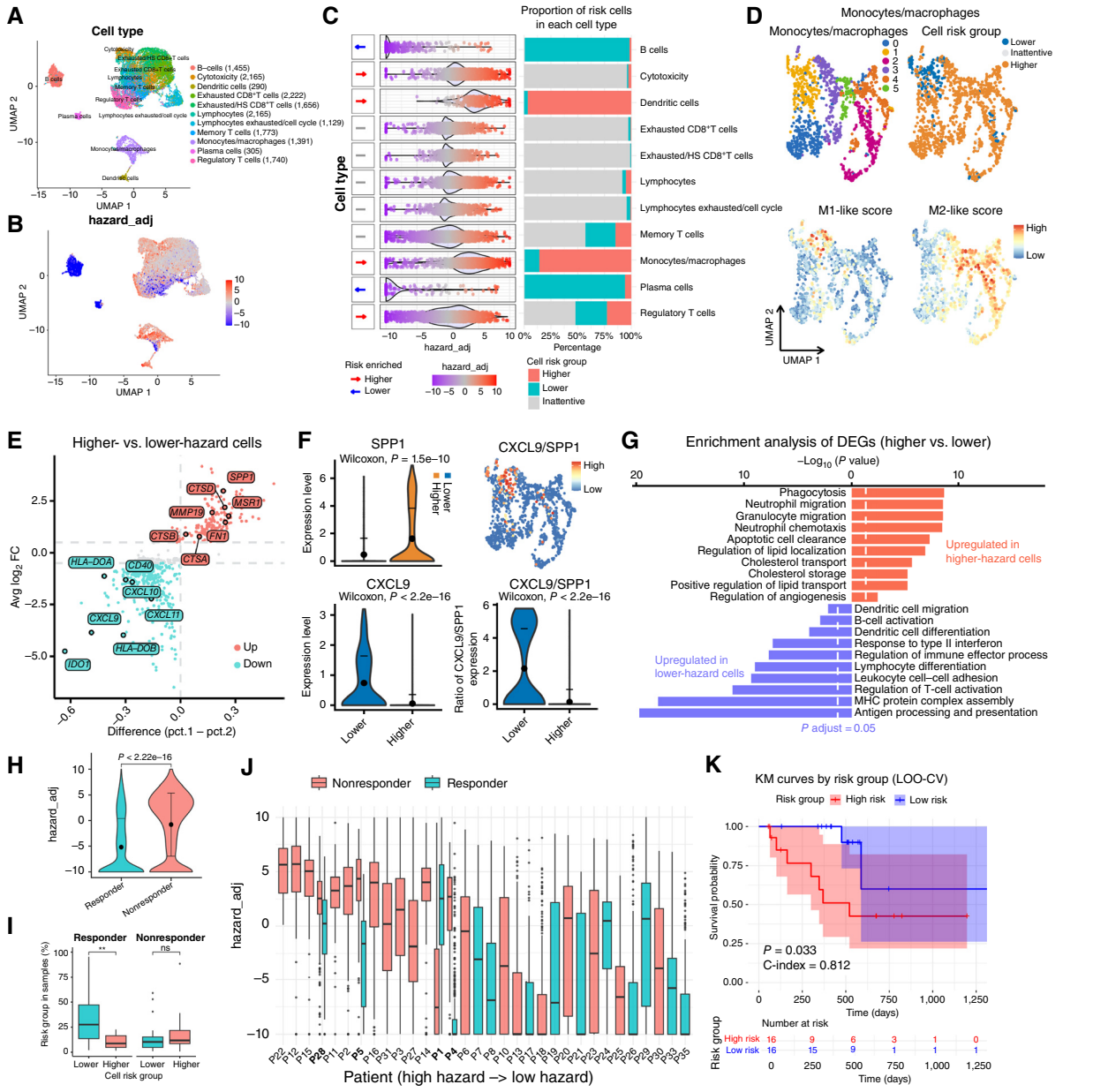

Figure 4. scSurvival analysis of a melanoma scRNA-seq cohort. **A,** UMAP visualization showing diverse cell types as defined in the original article (24, 25). Cell numbers are indicated in parentheses. **B,** UMAP showing adjusted hazard scores estimated by scSurvival for each cell. **C,** Distribution of adjusted hazard scores and proportions of higher-hazard, lower-hazard, and inattentive cells within each cell type. Red and blue arrows indicate cell types enriched for higher-hazard and lower-hazard populations, respectively. Arrow direction is determined by the position of the maximum density estimated by Kernel Density Estimation (KDE) relative to zero (threshold set to 1; only directions with an absolute maximum hazard greater than 1 are shown). **D,** UMAP visualization showing the distribution of six monocyte/macrophage clusters, higher-hazard and lower-hazard subsets, and enrichment score for M1-like and M2-like signatures. **E,** Scatter plot of DEGs between higher-hazard and lower-hazard monocytes/macrophages. Red dots represent genes upregulated in higher-hazard cells, blue dots represent genes upregulated in lower-hazard cells, and gray dots denote nonsignificant genes. **F,** Violin plots showing the expression of *SPP1*, *CXCL9*, and their expression ratio (*CXCL9/SPP1*) in higher-hazard versus lower-hazard monocytes/macrophages. UMAP plot with cells colored by the *CXCL9/SPP1* expression ratio. **G,** GO Biological Process enrichment analysis of DEGs between higher-hazard and lower-hazard monocytes/macrophages. **H,** Violin plot comparing adjusted hazard scores of cells between responders and nonresponders in the melanoma cohort. *P* value calculated using a one-sided Wilcoxon test. **I,** Boxplots comparing the proportions of higher-hazard versus lower-hazard groups within responders and nonresponders. *P* value calculated using a one-sided Wilcoxon test. **, $P < 0.01$; ns, not significant. **J,** Boxplots showing adjusted hazard scores of each patient, ranked from higher to lower patient-level hazard. **K,** Leave-one-out (LOO) CV at the patient level, evaluating the stratification ability of estimated patient-level hazard scores and reporting the overall C-index. *P* value calculated using the log-rank test. KM, Kaplan–Meier.

T cell–based survival model across multiple independent scRNA-seq melanoma datasets containing T-cell expression profiles (Supplementary Table S1).

In the T cell–only analysis, scSurvival identified a large number of key cells, with their attention and hazard scores shown in Fig. 5B and Supplementary Fig. S9A. Overall, cytotoxic

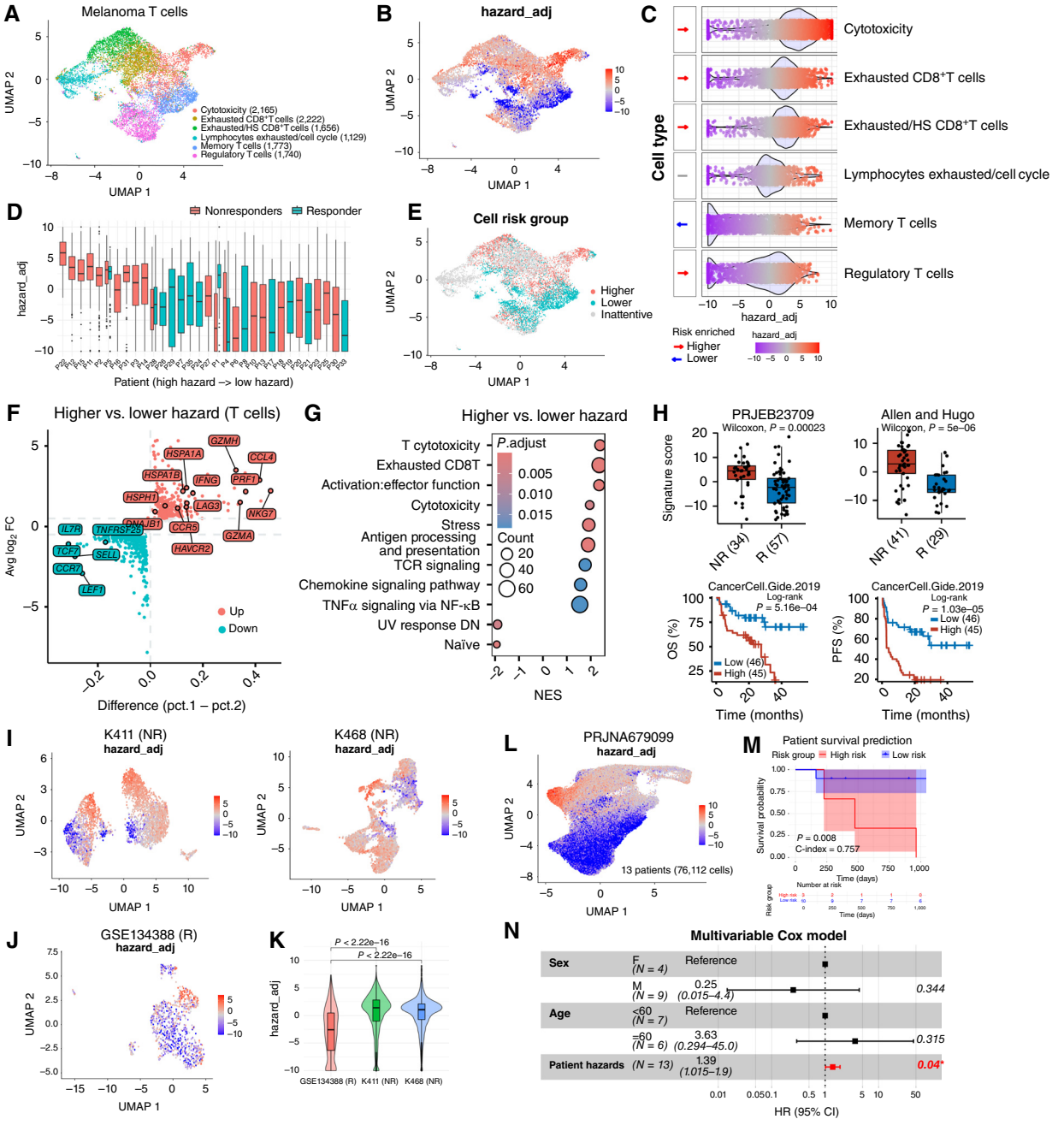

Figure 5. scSurvival analysis and prediction using T cells from a melanoma cohort. **A,** UMAP visualization generated using MVG2000, showing diverse T-cell subtypes. Cell numbers are indicated in parentheses. **B,** UMAP showing T-cell hazard scores estimated by scSurvival. **C,** Distribution of cell hazard scores across T-cell subtypes, with the overall risk trend indicated by red and blue arrows. **D,** Boxplot showing predicted hazard scores for each patient based on T cells, ranked by patient-level hazard scores. **E,** UMAP of all T cells colored by risk grouping. **F,** Scatter plot of DEGs between higher-hazard and lower-hazard T-cell groups. Red dots represent genes upregulated in higher-hazard cells, blue dots represent genes upregulated in lower-hazard cells, and gray dots denote nonsignificant genes. **G,** GSEA results comparing higher-hazard and lower-hazard T cells. **H,** A gene signature derived from DEGs between high-hazard versus low-hazard groups, evaluated in independent bulk RNA-seq melanoma datasets. Boxplots show signature scores in responder versus nonresponder groups across two independent datasets, with *P* values calculated using the Wilcoxon test. Kaplan–Meier survival curves compare patients with high versus low signature scores (median cutoff) in two independent datasets, with *P* values calculated using two-tailed log-rank tests. **I,** UMAP plots of predicted cell hazard scores in two independent nonresponder single-cell datasets using the trained scSurvival T-cell model. **J,** UMAP showing predicted cell hazard scores in an independent responder scRNA-seq sample using the trained scSurvival T-cell model. **K,** Violin plot comparing predicted hazard scores of cells from three new scRNA-seq samples. *P* value calculated using a one-sided Wilcoxon test. **L,** UMAP of predicted cell hazard scores of T cells from an independent melanoma cohort ($n = 13$ patients) using the trained scSurvival T-cell model. **M,** Risk stratification and C-index evaluation based on predicted patient-level hazard scores in this independent cohort. *P* value calculated using the log-rank test. **N,** Multivariable Cox regression analysis incorporating sex, age, and scSurvival-derived patient-level hazard scores. Data are presented as hazard ratio (HR) ± 95% confidence interval (CI) and *P* value was calculated using a Wald test. NES, normalized enrichment score; NR, nonresponder; R, responder; TCR, T-cell receptor.

T cells, exhausted CD8+ T cells, and regulatory T cells tended to be associated with longer survival, whereas memory T cells were linked to shorter survival (Fig. 5C). Notably, a subset of cytotoxic CD8+ T cells in this cohort expressed high levels of stress- and exhaustion-related markers (e.g., *TIGIT*, *HSPA1A*, and *HSPA1B*; Supplementary Fig. S9B), which may explain why these cells were associated with unfavorable outcomes in the Cox regression model (Supplementary Fig. S8A). Intrapatient cellular hazard scores remained largely consistent with immunotherapy response status (Fig. 5D). Furthermore, lower-hazard T cells were enriched in effective immune checkpoint blockade (ICB) responses, whereas higher-hazard T cells predominated in nonresponders (Fig. 5E; Supplementary Fig. S9C and S9D). Differential expression analysis revealed that lower-hazard T cells displayed elevated expression of genes associated with T-cell longevity and stem-like memory potential (34), such as *TCF7*, *CCR7*, and *IL7R* (Fig. 5F; Supplementary Fig. S9E; Supplementary Table S3). In contrast, higher-hazard T cells upregulated multiple immune-inhibitory receptors, including *HAVCR2* (TIM-3) and *LAG3*. Intriguingly, higher-hazard T cells also expressed a broad stress response program, with marked induction of heat shock genes such as *HSPA1A* and *HSPA1B* (Fig. 5F; Supplementary Fig. S9E). A recent pan-cancer study has described a distinct intratumoral T-cell subset, termed TSTR cells, characterized by pervasive heat shock gene expression and linked to poor clinical outcomes in immunotherapy (35). In line with these findings, gene set enrichment analysis (GSEA) revealed strong enrichment of cellular stress pathways, T-cell exhaustion signatures, and cytotoxic dysfunction programs in higher-hazard T cells (Fig. 5G; Supplementary Fig. S9F), reinforcing the notion that stress-adapted, dysfunctional T-cell states underlie poor therapeutic responses (35). Furthermore, the DEG-derived signature score was significantly higher in ICB nonresponders than in responders across multiple independent ICB datasets, and it effectively stratified both OS and PFS in the PRJEB23709 bulk RNA-seq melanoma immunotherapy cohort (Fig. 5H; ref. 36).

The trained scSurvival model can also be applied to predict patient survival outcomes in independent melanoma single-cell datasets. We first applied the model to three independent patients with melanoma treated with ICB, including two nonresponder samples (Fig. 5I; ref. 37) and one responder sample (Fig. 5J; ref. 38). The results showed that T cells in the responder sample had significantly lower hazard scores compared with those in the two nonresponder samples (Fig. 5K). We further validated the model on an independent melanoma cohort with survival outcomes (PRJNA679099; ref. 39), which includes 13 pretreatment samples comprising 76,112 cells (Supplementary Fig. S9G and S9H; Fig. 5L). The predicted patient hazard scores achieved a C-index of 0.757, and risk stratification (using optimal grouping) reached statistical significance ($P = 0.008$; Fig. 5M; Supplementary Fig. S9I). Univariate and multivariate Cox regression analyses based on the continuous patient hazard scores further confirmed that the predicted risks were significantly associated with survival outcomes ($P = 0.04$; Supplementary Fig. S9J; Fig. 5N). In addition, given that the training dataset included both pretreatment and posttreatment samples, we performed stratified analyses. The results showed that the features and signals identified by scSurvival were conserved across both pretreatment and posttreatment subsets (Supplementary Fig. S10A and S10B).

scSurvival Reveals Survival-Associated Cell Subpopulations in a Liver Cancer Cohort

We next applied scSurvival to a large liver cancer scRNA-seq atlas comprising 189 samples from 124 patients, including both tumor and TME cells with annotated cell types (40). Among these, survival information was available for 121 patients (Supplementary Fig. S11A; Supplementary Table S1). UMAP projection revealed diverse cell populations, including tumor/epithelial cells, multiple immune cell subtypes, and stromal components for a total of 1,092,172 cells (Fig. 6A). scSurvival computed cell-level hazard scores across all cells (Fig. 6B), which were further stratified into higher-hazard, lower-hazard, and inattentive groups via an attention-guided risk grouping strategy (Fig. 6C). Distinct cell types exhibited different distributions of hazard scores (Fig. 6D; Supplementary Fig. S11B). Consistent with previous results from the melanoma immunotherapy cohort analysis, scSurvival identified higher-hazard macrophages characterized by elevated *SPP1* and reduced *CXCL9* expression (Supplementary Fig. S11C). This expression pattern is also consistent with recent findings in liver cancer, in which SPP1+ TAMs promote tumor progression and stemness (41, 42).

In the tumor cell compartment, approximately one third of the epithelial cells were identified as higher hazard, with the remainder as lower hazard (Fig. 6D), highlighting substantial heterogeneity among tumor cells that contribute to patient prognosis. To dissect transcriptional differences, we compared gene expression profiles of higher- and lower-hazard tumor cells. As a result, 1,672 genes were upregulated and 997 genes were downregulated in higher-hazard tumor cells (Fig. 6E; Supplementary Table S4). Notably, the higher-hazard cells exhibited upregulation of genes linked to hypoxia (*HIF1A*), stemness (*PSCA* and *APLP1*), and epithelial–mesenchymal transition (EMT; *S100A4* and *SPP1*; ref. 43). GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses confirmed activation of EMT-related pathways, including wound healing, epithelium migration, and hypoxia-inducible factor-1 signaling pathway, in higher-hazard tumor cells (Fig. 6F; Supplementary Fig. S11D). Furthermore, GSEA revealed significant enrichment of EMT, TNFα signaling via NF-κB, hypoxia, and TGFβ signaling in the higher-hazard group, indicating a more invasive phenotype (Fig. 6G and H). In contrast, lower-hazard tumor cells showed elevated expression of liver-enriched metabolic genes such as *ARG1* and *ALDOB*. Correspondingly, pathway enrichment analysis highlighted activation of metabolic programs in the lower-hazard group (Fig. 6F and G; Supplementary Fig. S11D), suggesting that these cells exhibit heightened metabolic activity and a hepatocyte-like transcriptional profile. This observation aligns with previous reports that a subset of tumor cells can display a well-differentiated phenotype with hepatocyte-like features (44), which are typically associated with more favorable clinical outcomes in liver cancer.

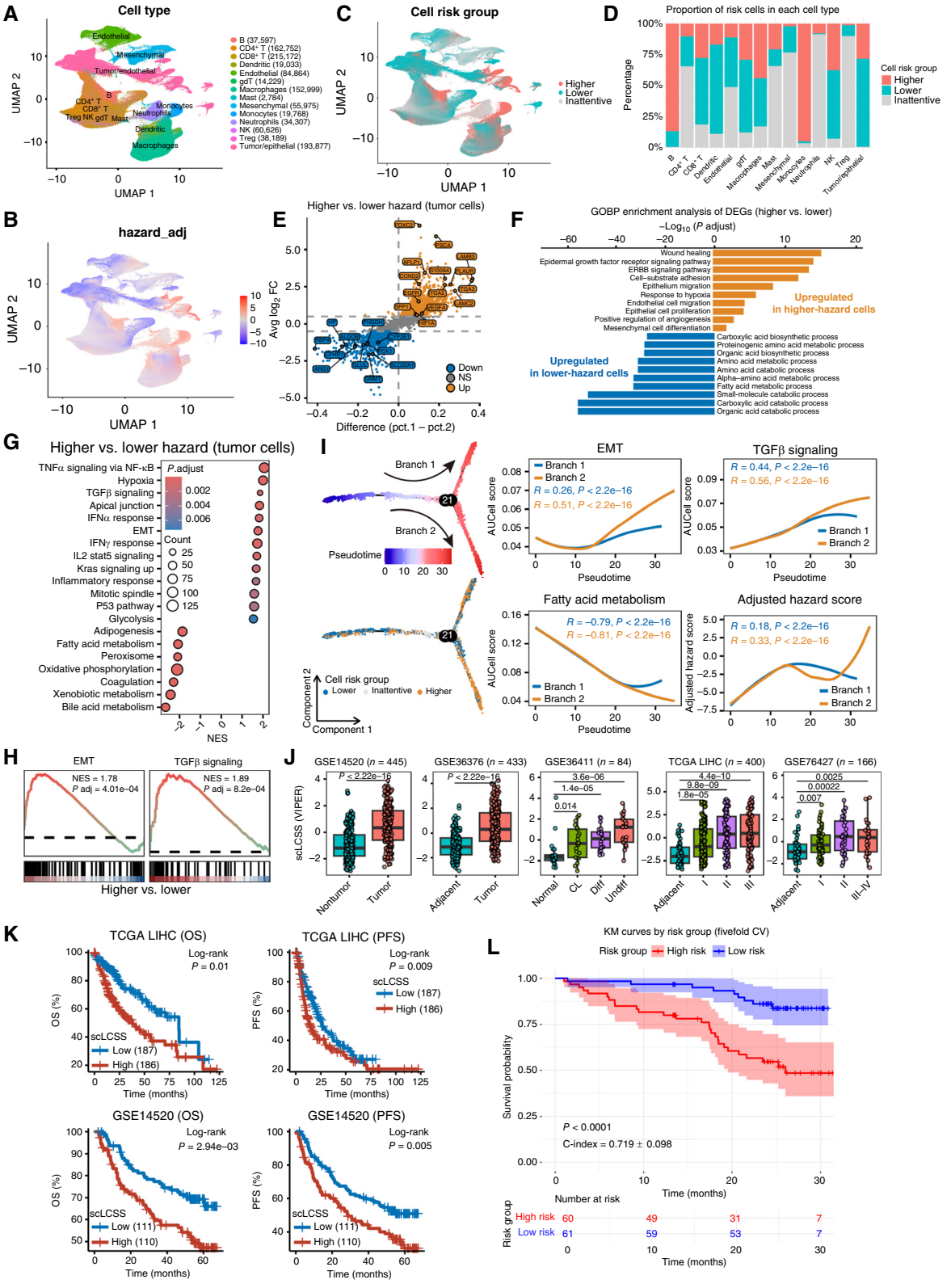

Figure 6. scSurvival analysis of a liver cancer single-cell cohort with survival data. **A,** UMAP visualization of a liver cancer scRNA-seq dataset colored by cell types with cell numbers indicated in parentheses. **B,** UMAP visualization of cell hazard scores estimated by scSurvival. **C,** UMAP visualization of cell risk groups, including higher-hazard, lower-hazard, and inattentive populations. **D,** Bar plots showing the proportions of higher-hazard, lower-hazard, and inattentive cells across cell types. **E,** Scatter plot of DEGs between high-hazard and low-hazard tumor cells. Orange dots indicate upregulated (up) genes, blue dots indicate downregulated (down) genes, and gray dots represent nonsignificant (NS) genes. (*continued on following page*)

Pseudotime analysis revealed a trajectory from lower-hazard tumor cells branching into two higher-hazard tumor states (Fig. 6I). Notably, tumor cells from branch 2 displayed elevated EMT activity and showed a stronger correlation with the adjusted hazard score (Fig. 6I), suggesting that this subpopulation is linked to greater invasiveness and poorer prognosis.

As a downstream analysis, we extended the scSurvival-identified higher-hazard and lower-hazard tumor cell states into bulk RNA-seq datasets to further assess their biological and clinical relevance. First, we combined the 100 most upregulated and 100 most downregulated DEGs into a unified single-cell liver cancer unfavorable survival signature (scLCSS; Supplementary Table S4). The scLCSS signature scores calculated on bulk RNA-seq datasets robustly distinguished tumor samples from nontumor or adjacent tissues, with scores increasing progressively with advancing tumor pathologic stage (Fig. 6J). More importantly, across four independent liver cancer cohorts with clinical follow-up, patients classified as scLCSS-high exhibited shorter OS and PFS than their scLCSS-low counterparts (Fig. 6K; Supplementary Fig. S11E). These findings highlight the ability of scSurvival to identify survival-associated cell subpopulations that are clinically relevant to prognosis.

Finally, we evaluated the predictive performance of scSurvival at the patient level. We conducted five-fold CV using patients from the liver cancer cohort. The predicted patient-level hazard scores in the testing folds achieved a C-index of 0.719 ± 0.098. Kaplan–Meier survival analysis, based on stratifying patients into high- and low-risk groups by the median of their predicted hazard score percentiles, further revealed that the high-risk group had significantly worse prognosis than the low-risk group ($P < 0.0001$; C-index = 0.719 ± 0.098; Fig. 6L). Consistently, univariate Cox regression showed that the scSurvival-derived hazard scores were significantly associated with survival ($P < 0.001$), outperforming other conventional immune features (Supplementary Fig. S8A). The achieved fivefold CV C-index (0.719) also exceeded those of benchmark models using cell type fractions, pseudo-bulk expression, or the *CXCL9/SPP1* ratio (Supplementary Fig. S8B). Together, these results demonstrate that the risk prediction model constructed by scSurvival using single-cell expression data can provide clinically meaningful survival predictions in real-world patient cohorts.

DISCUSSION

In this study, we develop scSurvival, a new framework that simultaneously constructs survival prediction models and identifies cell subpopulations associated with disease progression and patient outcomes. Unlike existing approaches that rely on bulk data as an intermediary for indirect association or that simply aggregate single-cell expression into pseudo-bulk profiles or cell type fractions, scSurvival adopts a multiple-instance learning strategy, modeling each patient's tumor profile as an ensemble of multiple cellular instances to enable cellular resolution. By introducing attention mechanisms and a multiple-instance Cox regression framework, scSurvival not only enables efficient and accurate hazard score prediction but also preserves cellular-level heterogeneity, allowing for the identification of key cell subpopulations closely associated with survival outcomes. Such survival-associated cell subpopulations identified by scSurvival could also be used to facilitate other forms of survival analysis such as pseudo-bulk analysis of those selected cell populations to build multigene risk scores. Our results demonstrate that scSurvival achieves robust performance across multiple simulated datasets and real clinical cancer cohorts. Notably, it successfully identified high-risk macrophage/monocyte subpopulations in the melanoma immunotherapy dataset and specific tumor cell subpopulations in the liver cancer cohort, providing valuable insights into the tumor immune microenvironment (TIME) and prognostic mechanisms. Moreover, a survival prediction model constructed using T cells from the melanoma cohort exhibited robust predictive accuracy when validated on independent datasets, further confirming the stability and broad applicability of scSurvival. Interestingly, our analysis revealed that a subset of $CD8^+$ cytotoxic T cells was unexpectedly associated with poorer prognosis in this cohort. Although counterintuitive, this likely reflects heterogeneous $CD8^+$ functional states, in which exhausted or stress-adapted subsets can override their canonical effector role (35). Similar context-dependent associations of $CD8^+$ T cells with unfavorable outcomes have been reported in certain cancer settings, underscoring the importance of considering their functional state rather than abundance alone when evaluating prognostic impact (24, 35). In addition, we observed context-dependent associations of B cells across tumor types. In melanoma, B cells were enriched in the lower-hazard group and frequently localized within TLS, consistent with their reported role in enhancing responses to ICB. By contrast, in liver cancer, B cells were enriched in the higher-hazard group, which may reflect the tolerogenic hepatic microenvironment in which B cells often acquire immunoregulatory phenotypes that suppress antitumor immunity. These findings highlight the heterogeneous and context-dependent roles of B cells in cancer prognosis (45, 46).

Figure 6. ***(Continued)*** **F,** GO Biological Process (GOBP) enrichment analysis of DEGs in high-hazard versus low-hazard tumor cells. Adjusted *P* values (*P*_adjust) calculated using the hypergeometric test and corrected for multiple testing using the FDR. **G,** GSEA of high-hazard tumor cells versus low-hazard tumor cells. **H,** GSEA enrichment curves of pathways altered in higher-hazard tumor cells versus lower-hazard tumor cells. **I,** Monocle 2 trajectory analysis of tumor cell differentiation, revealing two main divergent trajectories. Cells are colored by pseudotime (top) and by risk groups (bottom). Pearson correlation analysis between pseudotime scores and pathway activity or adjusted hazard score of cells. Smooth curves in each plot, generated using locally weighted smoothing method. **J,** Boxplots showing the distribution of scLCSS activity scores [inferred by the Virtual Inference of Protein Activity by Enriched Regulon analysis (VIPER) algorithm] across tumor and nontumor samples in five independent cohorts. **K,** Kaplan–Meier survival curves showing OS or PFS based on unfavorable survival signature scores in two liver cancer datasets. Patients were stratified into high- and low-risk groups according to the median risk scores. Statistical significance was determined using the two-tailed log-rank test. **L,** Five-fold CV at the patient level, evaluating the stratification ability of risk percentiles and reporting the mean C-index and SD. *P* value calculated using the log-rank test. LIHC, liver hepatocellular carcinoma; NES, normalized enrichment score; TCGA, The Cancer Genome Atlas; Treg, regulatory T cells.

Nevertheless, several limitations remain to be addressed in future work. First, although scSurvival incorporates mechanisms for batch effect correction, its robustness against extreme batch heterogeneity arising from different sequencing platforms or centers could be further improved. Future efforts may focus on enhancing batch effect modeling to facilitate more effective cross-batch data integration. Second, although the SE module currently serves to balance feature scales, its biological significance has not yet been fully explored. Future work could investigate the interpretability of SE weights or introduce alternative strategies for identifying key survival-related genes, thereby enhancing the biological transparency of the model. Third, although scSurvival supports the integration of clinical covariates at the patient level, we did not systematically evaluate how covariate adjustment influences the identification of risk-associated cell populations. In our initial evaluation, including covariates in scSurvival for the melanoma datasets produced similar results. Further studies are needed to dissect the interactions between covariate effects and intrinsic cellular risk features. Lastly, emerging spatial transcriptomics and proteomics technologies in cancer research provide rich molecular information coupled with spatial coordinates, offering an additional layer of biological context. The scSurvival framework could be extended to these spatial modalities by incorporating positional relationships among cells or spots, thereby enabling the identification of survival-associated, spatially co-localized patterns that extend beyond cell subpopulations.

In summary, scSurvival provides a scalable and systematic framework for survival prediction from cohort-level cancer single-cell data. It not only enables accurate prediction of patient survival outcomes but also provides a powerful tool for identifying survival-associated cell subpopulations to facilitate downstream mechanistic investigations at cellular resolution. By preserving cellular-level heterogeneity and leveraging attention-based modeling, scSurvival enhances our ability to link molecular and cellular features to patient outcomes. We believe that scSurvival will accelerate the broad adoption of cohort-level single-cell profiling for survival analysis and deepen our understanding of how specific cell subpopulations in the TME influence cancer outcomes, positioning it as a valuable tool for translational cancer research.

METHODS

scSurvival Framework

The input to scSurvival originates from a patient cohort (Fig. 1A), with a total of m patients. Each patient undergoes scRNA-seq, resulting in an individual gene expression matrix (Fig. 1B). The collection of all single-cell expression matrices was denoted as $X = \left\{X^{(i)}\right\}_{i=1}^{m}$, in which the expression matrix for patient i is represented as $X^{(i)} \in \mathbb{R}^{n_i \times p}$, comprising n_i cells and p genes. Each patient is also associated with survival information $Y = (T, E)$, in which $T = \left(t_1, t_2, \ldots, t_m\right) \in \mathbb{R}_+^m$ records the event times and $E = \left(e_1, e_2, \ldots, e_m\right) \in \{0,1\}^m$ indicates the event statuses, with $e_i = 1$ denoting the occurrence of the event (e.g., death) and $e_i = 0$ representing censoring (Fig. 1C; refs. 47, 48). To account for potential batch effects arising from different data sources, scSurvival also supports automatic batch effect correction, in which the batch labels corresponding to individual cells are denoted as $B = \left\{b^{(i)}\right\}_{i=1}^{m}$, in which $b^{(i)} \in \mathbb{R}^n$.

The scSurvival model primarily consists of two modules: a cell feature extraction module and a multiple-instance Cox regression module. The feature extraction module is a VAE-based generative model that learns unified cell embeddings across batches (Fig. 1D). In the multiple-instance Cox regression module, these cell embeddings are aggregated at the patient level through a multihead attention mechanism and subsequently passed into a scoring network to compute relative hazard scores and perform standard Cox regression (Fig. 1E). If patient-level covariates need to be considered, scSurvival also supports incorporating covariates simultaneously during Cox regression (Fig. 1G). In addition, we apply an entropy regularization (22) term on the attention weights to control the sparsity of attention and enable the identification of key cells. The overall loss function of the model is given by

$$\mathcal{L} = \mathcal{L}_{cox} + \lambda_1 \mathcal{L}_{VAE} + \lambda_2 \mathcal{L}_{Ent},$$

in which $\mathcal{L}_{cox}$, $\mathcal{L}_{VAE}$, and $\mathcal{L}_{Ent}$ denote the loss functions for Cox regression, VAE reconstruction, and entropy regularization, respectively, and λ_1 and λ_2 are hyperparameters balancing these components.

Ultimately, scSurvival is capable of providing survival predictions at the patient level, and based on the attention weights and the hazard scoring network, it can also identify key cells associated with higher or lower risk (with optional adjustment for covariate effects; Fig. 1I–K).

VAE Model Architecture

During the feature extraction phase, we train the model based on the complete set of single-cell expression data from all patients. In addition to standard quality control, preprocessing of the single-cell data before input to scSurvival includes total normalization within each cell, $\log_1 p$ transformation, and selection of highly variable genes (by default, scSurvival uses MVG2000 as the input; refs. 49, 50). Separating these preprocessing steps from the model facilitates reducing the model burden when making predictions on independent datasets. We assume that the preprocessed expression data are generated according to the following process:

$$p(z_c) = \mathcal{N}(z_c | 0, I),$$

$$p(x_{cg} | z_c, b_c) = ZIG(x_{cg} | \rho_{cg}, \mu_{cg}, \sigma_g^2),$$

$$\rho_{cg} = f_\rho^g(z_c, b_c; \varphi),$$

$$\mu_{cg} = f_\mu^g(z_c, b_c; \varphi).$$

Here, $z_c \in \mathbb{R}^{d_{model}}$ represents the low-dimensional latent state of cell c, in which d_{model} denotes the embedding dimension. Its prior distribution is assumed to follow the standard normal distribution, $\mathcal{N}(z_c | 0, I)$. The gene expression x_{cg} is modeled by a ZIG distribution (17, 18) with parameters ρ_{cg}, μ_{cg}, and σ_g^2, in which ρ_{cg} is the inflation probability, and μ_{cg} and σ_g^2 are the mean and variance of the nonzero Gaussian component, respectively. The ZIG distribution is capable of capturing technical dropouts in gene expression data and distinguishing them from biologically meaningful low expression. Its likelihood function is defined as

$$ZIG(x_{cg} | \rho_{cg}, \mu_{cg}, \sigma_g^2) = \rho_{cg} \delta(x_{cg}) + (1 - \rho_{cg}) \mathcal{N}(x_{cg} | \mu_{cg}, \sigma_g^2),$$

in which $\delta(x)$ is the Dirac delta function (51), which equals 1 at zero and 0 elsewhere. The parameters ρ_{cg} and μ_{cg} are computed by a parameterized decoder $f_\bullet(z_c, b_c; \varphi)$, which takes as input the latent state z_c and the corresponding one-hot encoded batch label b_c. The decoder is implemented as a fully connected neural network with parameters denoted by φ. The variance term σ_g^2 is a trainable, gene-specific parameter independent of the input.

According to the principle of VAE (arXiv 1312.6114), the posterior distribution of the latent state z_c is assumed to follow a Gaussian distribution of the form

$$q(z_c \mid x_c, b_c) = \mathcal{N}\left(z_c \mid \mu_\theta(x_c, b_c), \sigma_\theta^2(x_c, b_c)\right),$$

in which the parameters μ_θ and σ_θ^2 lie in $\mathbb{R}^{d_{model}}$ and are computed by an encoder that takes as input the expression vector x_c of cell c and its encoded batch label b_c. In practice, the latent state z_c is sampled via the following differentiable reparameterization trick:

$$z_c = \mu_\theta(x_c, b_c) + \sigma_\theta(x_c, b_c) \odot \varepsilon, \varepsilon \sim \mathcal{N}(0, I).$$

The encoder consists of three modules: LN, SE, and a feature extraction module. LN (19) normalizes the input during training. The SE module, short for squeeze-and-excitation module (20), is a lightweight attention mechanism defined as

$$\mathrm{SE}(x) = \mathrm{sigmoid}\left(W_e \cdot \mathrm{relu}\left(W_s \cdot x\right)\right) \odot x,$$

in which $W_s \in \mathbb{R}^{d \times p}$ projects the high-dimensional input features into a lower-dimensional space for global feature summarization, with d being the reduced dimension and p the original feature dimension (i.e., the number of genes). The matrix $W_s \in \mathbb{R}^{p \times d}$ maps the compressed features back to the original space to produce feature-wise reweighting coefficients. The SE module enables adaptive recalibration of features during training.

The feature extraction module within the encoder is implemented as a fully connected neural network, denoted as $E_\bullet(\bullet, \bullet; \theta)$ with parameters θ. It takes as input the concatenated recalibrated gene features and batch labels to compute the parameters of the latent posterior distribution:

$$\mu_\theta(x_c, b_c) = E_\mu\left(\mathrm{SE}\left(\mathrm{LN}(x_c)\right), b_c; \theta\right),$$

$$\sigma_\theta(x_c, b_c) = E_\sigma\left(\mathrm{SE}\left(\mathrm{LN}(x_c)\right), b_c; \theta\right).$$

VAE Reconstruction Loss and Rare Gene Weighting

Following the principle of β-VAE (52), we define the reconstruction loss using the negative evidence lower bound as

$$L_{VAE}(x_c) = -\sum_{g=1}^{p} \log ZIG\left(x_{cg}\right) + \beta D_{KL}\left(q(z_c \mid x_c, b_c) \,\|\, p(z_c)\right),$$

in which β is a hyperparameter controlling the strength of the Kullback-Leibler (KL) divergence (53) regularization. During training, we further employ the FreeBits technique (54) to control the influence of KL divergence and prevent feature collapse, by imposing a minimum KL divergence threshold τ independently on each latent dimension. In addition, to stabilize the variance estimates, we introduce a prior regularization (55) on σ_g^2 as

$$p\left(\sigma_g^2\right) = \mathrm{Gamma}\left(\sigma_g^2 \mid 2, 2\right),$$

and incorporate it into the VAE loss via maximum likelihood. The resulting overall loss function becomes

$$L_{VAE}(x_c) = -\sum_{g=1}^{p} \log ZIG\left(x_{cg}\right) + \beta D_{KL}\left(q(z_c \mid x_c, b_c) \,\|\, p(z_c)\right) - \lambda_\sigma \sum_{g=1}^{p} \log p\left(\sigma_g^2\right),$$

in which λ_σ is a hyperparameter controlling the strength of the variance prior regularization.

Moreover, to balance the expression levels across genes and prevent important lowly expressed rare genes from being dominated by highly expressed genes, we introduce gene-specific weights into the loss function. The gene weight is defined by

$$\bar{x}_g = \frac{1}{n} \sum_{c=1}^{n} x_{cg}, \text{for } g = 1, \ldots, p,$$

$$\tilde{x} = \mathrm{median}\left(\bar{x}_1, \bar{x}_2 .., \bar{x}_p\right),$$

$$\omega_g = \left(\frac{\bar{x}_g + \epsilon}{\tilde{x} + \epsilon}\right)^{-\alpha},$$

in which n is the total number of cells and ϵ is a small constant to avoid division by zero, and α is a hyperparameter controlling the strength of weighting differences.

Thus, the final VAE training loss becomes

$$L_{VAE}(x_c) = -\sum_{g=1}^{p} \omega_g \left(\log ZIG\left(x_{cg}\right) + \lambda_\sigma \log p\left(\sigma_g^2\right)\right) + \beta D_{KL}\left(q(z_c \mid x_c, b_c) \,\|\, p(z_c)\right).$$

The average loss over all cells is computed as

$$\mathcal{L}_{VAE} = \frac{1}{n} \sum_{c=1}^{n} L_{VAE}(x_c),$$

in which $n = \sum_{i=1}^{m} n_i$ denotes the total number of cells across all patients.

In addition to extracting cell features using ZIG-VAE, we also implement a standard VAE with mean-squared error as the reconstruction loss, given by

$$L_{G\text{-}VAE}(x_c) = -\sum_{g=1}^{p} \omega_g \left(x_{cg} - \mu_{cg}\right)^2 + \beta D_{KL}\left(q(z_c \mid x_c, b_c) \,\|\, p(z_c)\right),$$

and PCA (56) for feature extraction. These alternative methods can handle simpler scenarios and provide rapid performance validation in practical analyses.

Multi-Head Attention Aggregation Mechanism

We use the mean vector $\mu_\theta(x_c, b_c)$ output by the decoder as the stable embedding feature for each cell and subsequently use it as the input to the downstream model. For simplicity, we denote it as μ_c when there is no ambiguity. We then employ a multi-head attention mechanism (57) to aggregate the cell embeddings into patient-level features.

In each attention head, the cell embedding is first projected to a lower-dimensional space through a linear mapper:

$$h_c = W_h \mu_c + b_h,$$

in which $W_h \in \mathbb{R}^{d_h \times d_{model}}$ and $b_h \in \mathbb{R}^{d_h}$ are the weight matrix and bias vector, respectively, and $d_h = \frac{d_{model}}{n_h}$ is the reduced feature dimension, with n_h denoting the number of attention heads.

Next, we compute the gated attention coefficients (12) as

$$a_c = w_h^T \left(\tanh\left(U_h h_c\right) \odot \mathrm{sigmoid}\left(V_h h_c\right)\right),$$

in which $w_h \in \mathbb{R}^{d_h}$ and $U_h, V_h \in \mathbb{R}^{d_h \times d_h}$ are trainable parameters.

Subsequently, the dimension-reduced features and corresponding attention weights are grouped by patient, denoted as $\left\{\left(h_1^{(i)}, h_2^{(i)} ..., h_{n_i}^{(i)}\right)\right\}_{i=1}^{m}$ and $\left\{\left(a_1^{(i)}, a_2^{(i)} ..., a_{n_i}^{(i)}\right)\right\}_{i=1}^{m}$, respectively. For each patient, the attention coefficients are first normalized via the SoftMax function:

$$\tilde{a}_c^{(i)} = \frac{\exp\left(a_c^{(i)}\right)}{\sum_{c=1}^{n_i} \exp\left(a_c^{(i)}\right)}, \text{for } c = 1, .., n_i, \text{for } i = 1, .., m.$$

The patient-level feature is then obtained by the weighted aggregation of the cell features as

$$H^{(i)} = \sum_{c=1}^{n_i} \tilde{a}_c^{(i)} h_c^{(i)}, \text{for } i = 1, .., m.$$

Finally, the complete patient representation is formed by concatenating the outputs from all attention heads as

$$\mathrm{H}^{(i)} = \mathrm{Concat}\left(H_1^{(i)}, H_2^{(i)}, ..., H_{n_h}^{(i)}\right), \text{for } i = 1, .., m.$$

in which $H_j^{(i)}$ denotes the aggregated patient feature from the jth attention head.

Although the gated attention mechanism inherently tends to produce sparse attention, in practice, to better identify the most critical hazard cells, we further introduce entropy regularization (22) to explicitly control the sparsity of attention. First, the normalized information entropy for each patient is computed as

$$Ent^{(i)} = -\frac{1}{n_h}\sum_{j=1}^{n_h}\sum_{c=1}^{n_i}\frac{\tilde{a}_c^{(i,j)}\log\left(\tilde{a}_c^{(i,j)}\right)}{\log(n_i)},$$

in which $\tilde{a}_c^{(i,j)}$ denotes the normalized attention coefficient for cell c in patient i under the jth head. The entropy regularization loss is then defined as

$$\mathcal{L}_{Ent} = \max\left(-\frac{1}{m}\sum_{i=1}^{m}Ent^{(i)} - \gamma, 0\right),$$

in which γ is a threshold for the normalized entropy, such that no penalty is applied when the entropy decreases below γ.

Cox Regression

The aggregated patient-level features are subsequently used for standard Cox regression (47, 48). We define the hazard model as

$$h\left(t|\mathrm{H}^{(i)}\right) = h_0(t)\exp\left(s\left(\mathrm{H}^{(i)}\right)\right),$$

in which $h_0(t)$ denotes the baseline hazard and $s\left(\mathrm{H}^{(i)}\right)$ is the relative hazard score. The scoring function s is implemented as a simple two-layer neural network with the following structure:

$$o(h) = W_2\cdot \mathrm{dropout}_{0.5}\left(\mathrm{relu}\left(W_1\cdot h + b_1\right)\right) + b_2,$$

$$s(h) = \mathrm{clamp}\left(o(h), -10, 10\right),$$

in which $\mathrm{dropout}_{0.5}$ denotes dropout with a rate of 0.5 and clamp constrains the output within the range $[-10,10]$ to improve numerical stability.

The loss function is based on the partial likelihood of the Cox proportional hazards model, defined as

$$\mathcal{L}_{cox} = -\frac{1}{m_e}\sum_{i:e_i=1}\left(s\left(\mathrm{H}^{(i)}\right) - \log\sum_{i'\in R(t_i)}\exp\left(s\left(\mathrm{H}^{(i')}\right)\right)\right),$$

in which $m_e = \sum_{i=1}^{m} e_i$ denotes the total number of patients who experienced the event, e_i and t_i represent the event status and event time for patient i, respectively, and $R(t_i) = \left\{i' : t_{i'} \geq t_j\right\}$ denotes the risk set, comprising patients whose survival time is greater than or equal to t_i.

When incorporating patient-level covariates, we combine the relative hazard contributions from the covariates using a linear model. The hazard model is modified as

$$h\left(t|\mathrm{H}^{(i)}\right) = h_0(t)\exp\left(s\left(\mathrm{H}^{(i)}\right) + \eta^T X_{cov}^{(i)}\right),$$

in which $X_{cov}^{(i)}$ denotes the covariate vector for patient i and η represents the trainable weight vector associated with the covariates. The corresponding loss function becomes

$$\mathcal{L}_{cox} = -\frac{1}{m_e}\sum_{i:e_i=1}\left(s\left(\mathrm{H}^{(i)}\right) + \eta^T X_{cov}^{(i)} - \log\sum_{i'\in R(t_i)}\exp\left(s\left(\mathrm{H}^{(i')}\right) + \eta^T X_{cov}^{(i')}\right)\right).$$

Total Loss Function and Training Strategy

scSurvival trains the model by jointly optimizing the three loss components described above, with the total loss function defined as

$$\mathcal{L} = \mathcal{L}_{cox} + \lambda_1\mathcal{L}_{VAE} + \lambda_2\mathcal{L}_{Ent},$$

in which λ_1 and λ_2 are hyperparameters. In practice, to achieve stable cell embeddings, we adopt a pretraining followed by fine-tuning strategy. First, we pretrain the VAE by minimizing $\mathcal{L}_{VAE}$ alone using batch gradient descent. Subsequently, based on the pretrained VAE parameters, we fine-tune the model by jointly optimizing the overall loss $\mathcal{L}$ using the gradient surgery technique (58). The principle of gradient surgery is to apply different learning rates to different modules, thereby preserving the pretrained feature extractor's stability. During fine-tuning in scSurvival, we set the learning rate for the VAE module to be 1/10 of that for the multiple-instance Cox regression module.

This joint optimization strategy enables the model to learn cell representations that are adaptively aligned with survival outcomes, yielding more informative and task-specific embeddings than fixed features while simultaneously mitigating overcorrection of batch effects. Nevertheless, scSurvival also provides interfaces that allow users to directly input custom-fixed features if desired.

When handling extremely large-scale datasets (e.g., when the total number of cells exceeds one million), computing the Cox loss requires aggregating features from all cells, which in turn forces the VAE module, especially the decoder, to reconstruct all cells simultaneously. Because of the decoder's complexity and the reconstruction loss, this would result in constructing and storing extremely large computational graphs, causing significant memory overhead.

To address this, for ultra-large datasets, we adopt an alternating optimization strategy during fine-tuning to approximate joint training. Specifically, the steps are as follows:

1. Perform one gradient update step on $\lambda_1\mathcal{L}_{VAE}$. As only the VAE is updated in this step, we can employ mini-batch training as in the pretraining phase, which automatically releases the computational graph after each update and prevents memory accumulation. To further accelerate training, it is even possible to update using only a single batch. Empirically, this procedure is sufficient to maintain the stability of the pretrained model in large-scale settings.
2. Perform one full-batch gradient update on $\mathcal{L}_{cox} + \lambda_2\mathcal{L}_{Ent}$. Given the relative simplicity of the multiple-instance Cox module, full-batch updates remain manageable in terms of memory consumption.
3. Within each training iteration, execute steps 1 and 2 sequentially to implement alternating updates.

Cell Number Balancing Sampling

When the number of measured single cells varies significantly across patients, directly training on all data may lead to uneven distributions of attention coefficients among patients. For example, patients with a larger number of cells may have overall smaller attention values, introducing biases in cell-level comparisons. To address this issue, we introduce an optional cell number–balancing sampling strategy in the multiple-instance Cox regression module. Specifically, we first compute the average number of cells across all patients, denoted as $\bar{n}$. During each fine-tuning step, for patients with more than $\bar{n}$ cells, we perform uniform sampling without replacement; for patients with fewer than $\bar{n}$ cells, we perform uniform sampling with replacement. As a result, each patient retains exactly $\bar{n}$ cells for the computation of the Cox loss and entropy regularization loss, followed by gradient updates. As the sampling procedure is repeated at every training step, eventually, almost all cells participate in the training process without causing data resource waste.

Hyperparameter Settings

The threshold γ in the entropy regularization is the most critical hyperparameter in scSurvival as it directly controls the sparsity of the attention coefficients and thus influences the number of key cells retained. We recommend adjusting γ within the range of 0.5 to 1.

Other hyperparameters are kept consistent across all our experiments on both simulated and real datasets and can be categorized into model hyperparameters and training hyperparameters. The model hyperparameters include the input feature dimension, which is set to 2000 by default; in all analyses, we use MVG2000 genes selected

by Seurat (49) as input features; the latent feature dimension d_{model}, which is set to 128; the hyperparameters β and λ_σ for the VAE reconstruction loss, with default values of 1 and 0.1, respectively; the KL divergence threshold τ for FreeBits (54), which is set to 0.2; the gene weight scaling factor α, which is set to 0.2; the number of attention heads n_{head}, which is set to 8; and the loss weight hyperparameters λ_1 and λ_2, which are set to 0.01 and 1, respectively.

The training hyperparameters include a learning rate of 1×10^{-3}; 200 pretraining epochs; and a maximum of 500 fine-tuning epochs, with early stopping if no further improvement in the loss function or validation metrics is observed. We use the Adam optimizer during pretraining and switch to AdamW during fine-tuning, with a weight decay of 0.01. For small-scale datasets, the pretraining batch size is set to the maximum number of cells among all patients; if balanced sampling is enabled, the batch size is set to the sampling number (i.e., the average number of cells). For large-scale datasets, the batch size during VAE training can be appropriately increased depending on GPU memory availability, with a default value of 4,096.

Cell Importance and Cell Hazard

At the cell level, the attention coefficients reflect which cells play a critical role in risk estimation. Given the large number of cells, the values of the attention coefficients tend to be small after SoftMax normalization. Therefore, we define the importance of each cell based on the following form of attention coefficients:

$$\hat{a}_{c,j} = \mathrm{sigmoid}\left(a_{c,j}\right), \text{for } j = 1,..,n_{head},$$

$$\hat{a}_c = \max_{1\le j\le n_{head}}\left(\hat{a}_{c,j}\right).$$

in which $a_{c,j}$ denotes the unnormalized attention coefficient of cell c at the jth attention head. The sigmoid function rescales the raw attention values to the range (0–1), providing an interpretable measure of each cell's individual contribution to the model. Furthermore, to evaluate the relative hazard score at the cell level, we first align each cell's feature with the patient-level representation

$$\hat{h}_c = \mathrm{Concat}\left(h_{c,1}\hat{a}_{c,1},...,h_{c,j}\hat{a}_{c,j},...,h_{c,n_{head}}\hat{a}_{c,n_{head}}\right),$$

in which $h_{c,j}$ represents the linearly projected feature of cell c from the jth attention head and $\hat{a}_{c,j}$ selectively preserves the parts of the feature that make significant contributions to the model. Subsequently, the trained hazard scoring network is used to predict a relative hazard score for each cell as

$$s_c = s\left(\hat{h}_c\right).$$

Although the scoring function s is trained at the patient level, because of the enforced sparsity of the attention coefficients, the aggregated patient feature is effectively close to the features of a few key cells. As a result, the scoring function can still provide reasonable hazard estimates at the cell level.

This cell-level hazard score, modulated by the attention weights, effectively reflects the contribution-adjusted risk of each cell and is therefore referred to as the *adjusted hazard* (abbreviated as *hazard_adj*). Throughout this study, unless otherwise specified, all reported cell-level hazard scores refer to these attention-adjusted values.

Model Prediction

When there is no significant batch effect within the training dataset or between the training and testing datasets, we can ignore the batch variables during model training and directly make reasonable predictions on the test set. However, when substantial batch effects are present, prediction performance may deteriorate noticeably. The feature extraction module in scSurvival does not require clinical information from patients and can operate independently from the multiple-instance module. Therefore, scSurvival supports incorporating the test set into the feature extraction module during training, allowing the model to learn cross-dataset representations and enabling transfer learning, thereby improving prediction performance.

Evaluation Metrics

At the patient level, we typically use the C-index (59) to evaluate the model's predictive performance, defined as

$$\text{C-index} = \frac{\sum_{(i,i')\in\mathcal{C}} I\left(s_i > s_{i'}\right) + \frac{1}{2}\sum_{(i,i')\in\mathcal{C}} I\left(s_i = s_{i'}\right)}{|\mathcal{C}|},$$

in which $\mathcal{C} = \left\{(i,i') | i\neq i', e_i = 1, t_i < t_{i'}\right\}$ is the set of all comparable patient pairs, s_i denotes the risk score for patient i, t_i is the survival time for patient i, and $I(\cdot)$ is the indicator function.

When the test set contains a sufficiently large number of samples, directly computing the C-index on the test set is appropriate. However, when the test sample size is small, forming enough comparable pairs may be challenging.

This situation commonly arises during CV on small datasets. Therefore, we adopt the following procedure to compute evaluation metrics:

1. In each validation fold, predict the hazard scores for the test samples while also computing the hazard scores for the training samples.
2. To eliminate prediction bias across folds, transform the test sample hazard scores into quantiles relative to the training set risk score distribution.
3. Collect the quantile-transformed hazard scores across all validation folds and compute a unified C-index as the validation metric.

Note that all training steps, including the selection of most variable genes, are performed exclusively within the training set of each validation fold to avoid information leakage.

During training, we also introduce a variant of the C-index, termed the conditional C-index (CC-index), to monitor early stopping and improve generalization performance. The CC-index is defined as

$$\text{CC-index}(A) = \frac{\sum_{(i,i')\in\mathcal{C}'} I\left(s_i > s_{i'}\right) + \frac{1}{2}\sum_{(i,i')\in\mathcal{C}'} I\left(s_i = s_{i'}\right)}{|\mathcal{C}'|},$$

in which $\mathcal{C}' = \mathcal{C} - A\times A$ excludes all comparable pairs formed solely within set A. For evaluation, we merge the hazard scores from the validation and training sets and compute the CC-index by setting A as the training set. The CC-index evaluates not only the ranking accuracy within the validation set but also the relative ranking of validation samples within the combined dataset. As the CC-index incorporates comparable pairs between the validation and training sets, it remains applicable even for small validation sets. Note that the CC-index is only used for model selection during training and is not used for final performance reporting.

Additionally, we stratify patients based on the hazard score into high-risk and low-risk groups and evaluate the model risk stratification capability using Kaplan–Meier survival curves (60) and the log-rank test (61).

At the cell level, we classify cells into three groups based on their attention scores and hazard scores: inattentive, higher risk, and lower risk. Specifically, we first cluster cells into two groups based on their attention scores, labeling the low-attention group as inattentive. Within the high-attention group, we further divide cells into higher-risk and lower-risk groups based on whether their hazard scores are positive or negative. As ground-truth labels are available in the simulated datasets, we evaluate the performance of cell classification only on simulated data. We use confusion matrices, precision, recall, and F1-scores (62) to assess the agreement between the predicted cell group labels and the ground-truth labels.

Basic Simulation Data Setup

To simulate a cohort of single-cell datasets with survival-driven cell subpopulations, we first used Splatter (23) to generate a synthetic single-cell expression dataset containing 10,000 cells and 5,000 genes. The cells were simulated as belonging to three groups, with assignment probabilities set to 0.7, 0.15, and 0.15, respectively. The differential expression probabilities for the three groups were set to 0.2, 0.06, and 0.06, respectively. Among these, the two smaller groups were designated as cell subpopulations associated with good prognosis and poor prognosis, respectively, whereas the largest group was treated as background cells. Next, we generated a cohort of single-cell expression matrices along with corresponding survival times by sampling from this synthetic dataset. To ensure that the simulated survival outcomes were driven by the designed cell subpopulations, we modeled a dynamic progression: as survival time advances, the number of cells associated with poor prognosis decreases, whereas the number of cells associated with good prognosis increases; the number of background cells remains approximately constant throughout.

Suppose we simulate m samples with monotonically increasing survival times. For the i-th sample, the corresponding single-cell expression matrix is generated according to the following steps:

1. Set the sampling ratio $r = \frac{i}{m-1}$. Randomly sample a proportion r of cells from the good-prognosis group and a proportion $1 - r$ of cells from the poor-prognosis group;
2. Mix the sampled cells with the background cells and then randomly sample 1,000 cells from the combined pool to form the single-cell expression matrix for sample i;
3. Assign the event status to 1 and the survival time to i for sample i.

As the probabilities of the two survival-driving cell groups are initially balanced, their total cell numbers are approximately equal. Through the above sampling procedure, the total number of mixed cells remains nearly constant for each sample, whereas the proportion of good-prognosis and poor-prognosis cells changes progressively, and the proportion of background cells remains stable. In addition, for each sample, we simulated censoring with a probability of 0.1. If censoring occurs, we set the event status to 0 and randomly advance the survival time (i.e., the censoring time). Based on this process, we generated a cohort of 100 samples with associated single-cell expression matrices and survival outcomes for evaluating the performance of scSurvival in identifying survival-driving cells and predicting survival outcomes.

To further validate the model's ability to integrate single-cell data across different batches, we also simulated patients originating from different batches. Specifically, using Splatter, we simulated two batches of data containing 6,000 and 4,000 cells, respectively, each with 5,000 genes. Each batch contained three cell clusters corresponding to good prognosis, poor prognosis, and irrelevant (background) groups. We then applied the same sampling strategy as above to generate 50 samples from each batch, each with single-cell expression matrices, survival times, and censoring statuses.

Extended Simulation Setup

To comprehensively evaluate the robustness and generalizability of scSurvival under various biological and technical conditions, we designed six types of extended simulation scenarios (Sim.1–Sim.6) based on the basic settings. Each scenario introduced a specific perturbation mechanism, and multiple parameter settings were set per scenario to simulate varying degrees of complexity and noise.

1. Risk cell percentage (Sim.1): We varied the proportions of good.survival and bad.survival cells in the single-cell reference dataset used for patient sampling, thereby generating datasets with different sizes of ground-truth risk cell populations. This setting was used to assess the stability of scSurvival when identifying rare risk subpopulations.
2. Batch imbalance (Sim.2): To simulate risk signal imbalance across batches, we generated two batches of single-cell data and downsampled good.survival or bad.survival cells asymmetrically between batches. The downsampling ratio controlled the degree of imbalance, with extreme settings in which one risk cell type was present in only a single batch. This setting tested the model's ability to handle risk feature shift across batches.
3. Patient outcome distribution (Sim.3): To simulate more realistic, noise-prone outcome distributions, we modeled patient survival times using an accelerated failure time (AFT) framework with a Weibull baseline distribution (63). The proportion of risk-associated cells, denoted as s, was used as the primary covariate. Specifically, the survival time T for each patient was generated according to the following formula:

$$T = \exp\left(-\beta \cdot \frac{s}{\mathrm{IQR}(s)}\right) \cdot \left[\alpha \cdot (-\ln U)^{\frac{1}{\gamma}}\right],$$

in which $\mathrm{IQR}(s)$ denotes the IQR of s across all patients, used to normalize the effect scale across different settings; $\beta = \ln(TR)$, in which TR is the time ratio, defined as the median survival time ratio between high- and low-risk patients, indicating the extent to which a covariate influences expected survival time; α is the scale parameter of the survival distribution, fixed at 10; γ is the shape parameter, set to 1.5 to simulate higher risk in later stages of disease progression, such as cancer; and U is a random variable sampled from $\mathrm{Uniform}(0.2, 0.8)$ to introduce stochasticity into survival times.

We tested three levels of effect size, corresponding to TR values of 1.3, 2, and 3, to reflect varying degrees of biological signal strength. Additionally, we included a noiseless idealized setting ($TR = \inf$), in which U was fixed at 0.5 to yield deterministic survival times. Moreover, to simulate event censoring, we introduced random censoring under three censoring probabilities: 0.2, 0.5, and 0.7. When censoring occurred, the censoring time was uniformly sampled between 0 and the simulated survival time generated by the AFT–Weibull model.

4. Cell composition heterogeneity (Sim.4): To evaluate the robustness of scSurvival under complex cellular compositions, we simulated more complex celltype compositions for each patient. Specifically, the proportions of background cell types were randomly varied across patients to introduce heterogeneity and noise. The degree of cell composition heterogeneity was controlled by adjusting the number of distinct background cell types, allowing us to assess whether scSurvival could consistently identify true risk-associated cells under varying levels of complexity.
5. Patient group heterogeneity (Sim.5): Beyond heterogeneity arising from cell composition, we also considered interpatient variability in how individuals respond to the same biological signals. For instance, elderly patients may exhibit stronger clinical responses to the same level of risk signal compared with younger patients. To simulate this group-wise covariate effect, we extended the AFT–Weibull model to incorporate a binary grouping variable as

$$T = \exp\left(-(\beta + g\delta) \cdot \frac{s}{\mathrm{IQR}(s)}\right) \cdot \left[\alpha \cdot (-\ln U)^{\frac{1}{\gamma}}\right],$$

in which $\delta \in \{0,1\}$ indicates whether the patient belongs to a high-response group (e.g., elderly) and g represents the magnitude of effect shift between the two groups. We varied both the proportion of patients assigned to the high-response group and the value of g to systematically evaluate the ability of scSurvival to model patient-level heterogeneity, and accordingly, this simulation scenario was analyzed using the covariate-aware mode of scSurvival.

6. Risk subcluster heterogeneity (Sim.6). To highlight the advantage of scSurvival in performing survival analysis at true single-cell resolution, we introduced a challenging simulation scenario in which patient survival time originates only from a small subpopulation

embedded within a larger cell cluster. We varied two key parameters: (i) the proportion of ground-truth risk-driving cells within the larger cluster and (ii) the degree of separation between the risk subpopulation and the rest of the cluster, achieved by adjusting the log fold change of DEGs in Splatter (23). This setting was designed to assess whether scSurvival can still effectively model survival outcomes and accurately identify relevant subpopulations even in cases in which conventional clustering algorithms might fail because of the rarity or subtlety of the risk signal.

Finally, based on the simulation approach described in Sim.3, we fixed the censoring rate at 0.3 and generated single-cell cohorts with varying numbers of patients and cells per patient. This allowed us to evaluate, under different effect sizes, how many patients are required for scSurvival to reliably identify risk-driving cell subpopulations.

Besides, for benchmarking and visualization purposes, we selected one representative setting from each simulation category, reflecting moderately challenging but nonextreme scenarios (Fig. 2G):

- Sim.1: risk cell proportion = 3%;
- Sim.2: downsampling rate = 0.2;
- Sim.3: censoring rate = 0.7 and TR = 2;
- Sim.4: number of background clusters = 5;
- Sim.5: group effect $g=3$ and older group proportion = 0.5;
- Sim.6: subcluster proportion = 0.3 and separation = 0.35.

Benchmark Strategies

To systematically evaluate the performance of scSurvival, we designed two complementary benchmarking strategies. The first benchmarking strategy focused on identifying the ground-truth risk-driving cell subpopulations in the simulated datasets. To isolate the impact of feature extraction, we fixed the downstream multi-instance Cox regression module and replaced only the upstream feature encoder. Specifically, we compared the default (ZIG-VAE with alternative feature representations, including PCA, NMF (implemented via sklearn.decomposition.NMF; ref. 64), and top 2000 highly variable genes (HVG2000) identified by Scanpy (RRID: SCR_018139; ref. 50), which were used directly without further dimensionality reduction. Note that all methods other than ZIG-VAE did not include any explicit batch effect correction modules. The resulting cell-level features were passed into the same multi-instance Cox model, in which risk-relevant cells were identified by jointly analyzing attention weights and estimated cell-level hazard scores. Model performance was evaluated by comparing predicted risk cell groupings against the known ground-truth labels using macro-averaged precision, recall, and F1-score under a three-class setting (good.survival, bad.survival, and other).

The second benchmarking strategy aimed to evaluate patient-level survival prediction. We constructed Cox regression models using either cell type proportions or pseudo-bulk expression as input features, serving as baseline methods for comparison with scSurvival. All models were implemented using sksurv.linear_model.CoxnetSurvivalAnalysis (65) and evaluated via nested CV, with a fivefold outer loop to assess model performance and an inner loop for tuning the regularization parameter (α). For the cell type proportion model, the input consisted of the relative abundance of major cell types per patient. To approximate ridge regression and improve stability under feature collinearity, the l1_ratio was set to 0.1 in simulation datasets and 1e−3 in real-world data. Cell types in simulation were derived using the Leiden clustering algorithm (66). For the pseudo-bulk model, the input was the average gene expression profile per patient, and the l1_ratio was set to 1 (lasso regression) to facilitate feature selection. In real-data analysis, we additionally constructed a univariate Cox model based on the CXCL9/SPP1 expression ratio, a widely used immune-related prognostic marker. All models were evaluated by calculating the C-index on the outer CV loop to assess risk prediction accuracy.

GSEA

Differential expression analysis between higher-hazard and lower-hazard cells was carried out with significance thresholds of $|\log_2 \text{fold change}| > 0.5$ and adjusted $P < 0.05$. As previously described (67), we conducted GO, KEGG, and rank-based GSEA using the R package clusterProfiler (v4.7.1; RRID: SCR_013042). For bulk transcriptomic data, we employed the single-sample GSEA method with the R package GSVA. For scRNA-seq data, the AUCell algorithm was employed to mitigate the impact of dropouts and inherent technical variability. Gene sets used in this study, including Hallmark (v7.2), GO (v7.2), and KEGG (v7.2), were retrieved from the Molecular Signatures Database (v7.2; RRID: SCR_016863).

Kaplan–Meier Survival Curve Comparison

As previously described (9), we conducted ordinary survival analyses to assess the clinical relevance of the scSurvival-derived gene signatures. The resulting survival signature comprised the 100 most upregulated and the 100 most downregulated DEGs from higher-hazard versus lower-hazard cells. Signature scores were computed using the Virtual Inference of Protein Activity by Enriched Regulon analysis algorithm (68), which integrates both the upregulated and downregulated scSurvival-derived gene sets. Based on the median value of the signature scores, samples were dichotomized into high-score and low-score groups. Kaplan–Meier survival curves were generated for these groups, and statistical significance was determined using the log-rank test.

Public Dataset Collection and Processing

In this study, we curated four publicly available human scRNA-seq datasets and one spatial transcriptomic dataset for both discovery and validation (Supplementary Table S1). For the PRJNA679099 scRNA-seq dataset, raw FASTQ files were downloaded from the NCBI BioProject database. Raw gene expression matrices were generated using CellRanger (v8.0.1; RRID: SCR_017344) with the refdata-gex-GRCh38-2020-A reference genome, following the standard 10x Genomics pipeline with default parameters unless otherwise specified. The expression matrices for the other datasets were downloaded from their respective public repositories.

After importing raw counts, we removed low-quality cells (fewer than 250 detected genes, fewer than 500 total transcripts, or more than 20% mitochondrial reads). Seurat (v 5.1.0; RRID: SCR_016341) was then applied to perform normalization, identification of variably expressed genes, and PCA. A shared nearest-neighbor graph built from the first 10 principal components was used to delineate communities. Finally, all cells were visualized in two dimensions with UMAP.

Human bulk transcriptomic datasets were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) and UCSC Xena database (https://xena.ucsc.edu/), and the detailed information is summarized in Supplementary Table S1. For the PRJEB23709 melanoma bulk RNA-seq dataset, raw FASTQ files were downloaded from the NCBI BioProject database. As described in the official documentation (69), raw gene expression matrices were generated using the Salmon pipeline (v1.4.0; RRID: SCR_017036) following the recommended preprocessing workflow (https://github.com/COMBINE-lab/salmon). The expression matrices for the other bulk transcriptomic datasets were downloaded from their respective public repositories. For all bulk RNA-seq data, the transcript per million value was calculated.

Main scRNA-Seq Dataset Descriptions

In the GSE120575 melanoma immunotherapy cohort, we analyzed 48 tumor samples from 32 patients, encompassing a total of 16,291 $CD45^+$ immune cells profiled using an optimized full-length Smart-seq2 protocol. Cell type annotations were derived from the original

study (24). Based on radiologic assessments of immunotherapy response, the samples were classified into 17 responders and 31 nonresponders (24). In the present study, this dataset was used as both a training set and discovery cohort for identifying survival-associated immune cell subpopulations and constructing predictive transcriptional signatures. The processed expression matrix was downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120575), and only protein-coding genes were retained for all analyses performed in this study. The detailed preprocessing steps are described in the Methods section. Besides, we collected one spatial transcriptomic dataset (28) and three bulk RNA-seq datasets, including PRJEB23709 (36), GSE78220 (70), and phs000452.v3.p1 (71), to further evaluate the robustness of the derived signature.

To assess the risk landscape of $CD8^+$ T cells in melanoma under ICB, we analyzed scRNA-seq datasets from three patients with annotated treatment outcomes. Two nonresponding metastatic melanoma samples (K411 and K468) were obtained from GSE159252 (37). Quality control followed the original publication's parameters in Seurat by excluding cells with fewer than three counts, fewer than 400 detected features, or more than 10% mitochondrial content. $CD8^+$ T-cell subsets were extracted for downstream analysis, yielding 3,737 cells from K411 and 4,878 cells from K468. For the responding case, we accessed the processed expression matrix and cell type annotations of sample GSE134388 (38) from the TISCH24 portal (https://tisch.compbio.cn/home/) and similarly extracted the $CD8^+$ T-cell subset. Hazard-adjusted values were projected onto UMAP for visualization and compared across patients.

In addition, the PRJNA679099 melanoma scRNA-seq dataset with available survival information was used as an independent validation cohort (39). The PRJNA679099 dataset comprises scRNA-seq data from 13 immunotherapy-naïve patients with cutaneous melanoma. In this study, researchers performed 10x Genomics scRNA-seq on sorted immune cells from resected tumors and peripheral blood, profiling tumor-infiltrating lymphocytes. In the present work, this dataset was used as an independent validation cohort for evaluating the performance of the prognostic model. Raw FASTQ files were downloaded from the NCBI BioProject database (PRJNA679099), and upstream processing was performed to generate expression matrices (see "Methods" section for details).

Furthermore, we included the GSE221553 melanoma scRNA-seq dataset (33) as an another independent validation cohort. This dataset comprises longitudinal tumor samples from 13 patients with metastatic melanoma treated with tumor-infiltrating lymphocyte adoptive cell therapy. The study used CD45-sorted scRNA-seq to interrogate the TME across treatment time points. In our analysis, we downloaded raw count matrix from GEO, performed quality control, normalization, and cell type annotation, and then applied our scSurvival pipeline to validate whether the molecular signatures, hazard-associated cell populations, and pathway enrichments identified in our training melanoma dataset are recapitulated in this independent scRNA-seq dataset.

The PRJCA007744 dataset comprises scRNA-seq data from 189 samples collected from 124 patients with liver cancer (40). This comprehensive dataset was generated using the 10x Genomics platform and includes more than 1 million cells, encompassing various liver cancer subtypes such as hepatocellular carcinoma, intrahepatic cholangiocarcinoma, and combined hepatocellular cholangiocarcinoma. In the corresponding study, researchers performed scRNA-seq analysis to investigate the heterogeneity of the TIME in liver cancer (40). In the present study, this dataset was used as both a training set and discovery cohort for identifying survival-associated tumor cell subpopulations and constructing predictive transcriptional signatures. The processed raw count matrix was obtained directly from the authors (40), and we removed single-cell data from three patients without available survival information. Only protein-coding genes were retained, and ribosomal genes (RPL and RPS) and mitochondrial genes (MT-) were excluded. The detailed preprocessing steps are described in the "Methods" section. Because of the lack of publicly available scRNA-seq datasets with survival annotations, we additionally collected eight bulk transcriptomic datasets as external validation cohorts (Supplementary Table S1), including GSE36411 (72), GSE14520 (73), TCGA-LIHC (74), GSE76427 (73), GSE40873 (75), and GSE116174 (76).

Data Availability

The public datasets analyzed in this study are described in the "Methods" section, with corresponding accession numbers provided therein. For convenience, all accession numbers and dataset details are also compiled in Supplementary Table S1. The open-source scSurvival program and its tutorials are freely available at GitHub (https://github.com/cliffren/scSurvival), Zenodo (https://doi.org/10.5281/zenodo.15399777), and Code Ocean (https://codeocean.com/capsule/7948185/tree/v1).

Authors' Disclosures

G.B. Mills reports personal fees and nonfinancial support from Amphista Therapeutics, Astex Pharmaceuticals, AstraZeneca, BlueDot, Ellipses Pharmaceuticals, ImmunoMET Therapeutics, Ionis Pharmaceuticals, Leapfrog Bio, Bruker/NanoString Technologies, NeoPhore, Nerviano Medical Sciences, Nuvectis Pharma, Pangea Pharmaceuticals, PDX Pharmaceuticals, Qureator, RyboDyn, SignalChem Lifesciences, Tarveda Therapeutics, Turbine, and Zentalis Pharmaceuticals; ownership of BlueDot stock, Catena Pharmaceuticals stock, ImmunoMet Therapeutics stock, Nuvectis Pharma stock, RyboDyn stock, SignalChem Lifesciences stock, Tarveda Therapeutics stock, and Turbine stock; other support from Adelson Medical Research Foundation, Breast Cancer Research Foundation, Komen Research Foundation, Ovarian Cancer Research Foundation, and Prospect Creek Foundation; other support from AstraZeneca (sponsored research), Zentalis Pharmaceuticals (sponsored research), NanoString (sponsored research), and Ionis Pharmaceuticals (sponsored research); other support from AstraZeneca (clinical trial support), Genentech (clinical trial support), GSK (clinical trial support), and Eli Lilly and Company (clinical trial support) during the conduct of the study; grants, personal fees, and nonfinancial support from Amphista Therapeutics, Astex Pharmaceuticals, AstraZeneca, BlueDot, Ellipses Pharmaceuticals, ImmunoMET Therapeutics, Ionis Pharmaceuticals, Leapfrog Bio, Bruker/NanoString, NeoPhore, Nerviano Medical Sciences, Nuvectis Pharma, Pangea Pharmaceuticals, PDX Pharmaceuticals, Qureator, RyboDyn, SignalChem Lifesciences, Tarveda Therapeutics, Turbine, and Zentalis Pharmaceuticals; ownership of BlueDot stock, Catena Pharmaceuticals stock, ImmunoMet Therapeutics stock, Nuvectis Pharma stock, RyboDyn stock, SignalChem Lifesciences stock, Tarveda Therapeutics stock, and Turbine stock, other support from Adelson Medical Research Foundation, Breast Cancer Research Foundation, Komen Research Foundation, Ovarian Cancer Research Foundation, and Prospect Creek Foundation, other support from AstraZeneca (sponsored research), Zentalis Pharmaceuticals (sponsored research), NanoString (sponsored research), and Ionis Pharmaceuticals (sponsored research), and other support from AstraZeneca (clinical trial support), Genentech (clinical trial support), GSK (clinical trial support), and Eli Lilly and Company (clinical trial support) outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

T. Ren: Data curation, software, formal analysis, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing. **F. Zhao:** Data curation, formal analysis, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing. **C. Chen:** Data curation,

writing–review and editing. **L. Zhou:** Data curation. **L.-Y. Wu:** Methodology, writing–review and editing. **G.B. Mills:** Investigation, methodology, writing–review and editing. **L.M. Coussens:** Investigation, methodology, writing–review and editing. **Z. Xia:** Conceptualization, resources, data curation, supervision, funding acquisition, investigation, methodology, writing–original draft, project administration, writing–review and editing.

Acknowledgments

This work was supported by the following funding sources: the Department of Defense Prostate Cancer Data Science Award HT94252410551, NIH R01GM147365 and R01CA283171, and the Silver Family Innovation Foundation Award (to Z. Xia); NIH U01CA253472, U01CA281902, and U24CA264128 (to G.B. Mills); and the Susan G. Komen Foundation, the National Foundation for Cancer Research, and Hildegard Lamfrom Endowed Chair in Basic Research (to L.M. Coussens). The research reported in this publication used computational infrastructure supported by the Office of Research Infrastructure Programs, Office of the Director, of the NIH under Award Number S10OD034224. We acknowledge the The Cancer Genome Atlas Research Network for generating, curating, and providing access to the liver hepatocellular carcinoma dataset used in this study. We also acknowledge the use of publicly available scRNA-seq datasets deposited in the GEO and the Sequence Read Archive. We thank the investigators who generated and shared these valuable datasets. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

Received June 3, 2025; revised October 8, 2025; accepted December 24, 2025; posted first April 21, 2026.

REFERENCES

1. Zhou J-G, Zhao H-T, Jin S-H, Tian X, Ma H. Identification of a RNA-seq-based signature to improve prognostics for uterine sarcoma. Gynecol Oncol 2019;155:499–507.
2. Gu Y, Lu L, Wu L, Chen H, Zhu W, He Y. Identification of prognostic genes in kidney renal clear cell carcinoma by RNA-seq data analysis. Mol Med Rep 2017;15:1661–7.
3. Raman P, Zimmerman S, Rathi KS, de Torrenté L, Sarmady M, Wu C, et al. A comparison of survival analysis methods for cancer gene expression RNA-sequencing data. Cancer Genet 2019;235–236:1–12.
4. Huang Z, Johnson TS, Han Z, Helm B, Cao S, Zhang C, et al. Deep learning-based cancer survival prognosis from RNA-seq data: approaches and evaluations. BMC Med Genomics 2020;13(Suppl 5):41.
5. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009;6:377–82.
6. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc 2010;5:516–35.
7. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. Nat Rev Nephrol 2018;14:479–92.
8. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell 2015;58:610–20.
9. Sun D, Guan X, Moran AE, Wu L-Y, Qian DZ, Schedin P, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. Nat Biotechnol 2022;40:527–38.
10. Mizukoshi C, Kojima Y, Hayashi S, Abe K, Kasugai D, Shimamura T. scSurv: a deep generative model for single-cell survival analysis. Bioinformatics 2026;42:btaf671.
11. Johnson TS, Yu CY, Huang Z, Xu S, Wang T, Dong C, et al. Diagnostic Evidence GAuge of Single cells (DEGAS): a flexible deep transfer learning framework for prioritizing cells in relation to disease. Genome Med 2022;14:11.
12. Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. In: Jennifer D, Andreas K, editors. Proceedings of the 35th International Conference on Machine Learning. Volume 80. Proceedings of Machine Learning Research. Long Beach (CA): PMLR; 2018. p. 2127–36.
13. Maron O, Lozano-Pérez T. A framework for multiple-instance learning. Advances in Neural Information Processing Systems. 1997. Available from: https://proceedings.neurips.cc/paper_files/paper/1997/file/82965d4ed8150294d4330ace00821d77-Paper.pdf
14. Carbonneau M-A, Cheplygina V, Granger E, Gagnon G. Multiple instance learning: a survey of problem characteristics and applications. Pattern Recognition 2018;77:329–53.
15. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods 2018;15:1053–8.
16. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Mol Syst Biol 2021;17:e9620.
17. Pierson E, Yau C. ZIFA: dimensionality reduction for zero-inflated single-cell gene expression analysis. Genome Biol 2015;16:241.
18. Ghosh SK, Mukhopadhyay P, Lu J-CJ. Bayesian analysis of zero-inflated regression models. J Stat Plann Inference 2006;136:1360–75.
19. Xu J, Sun X, Zhang Z, Zhao G, Lin J. Understanding and improving layer normalization. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook (NY): Curran Associates Inc.; 2019. p. 394.
20. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell 2020;42:2011–23.
21. Tukey JW. Use of many covariates in clinical trials. Int Stat Rev 1991;59:123–37.
22. Yves G, Yoshua B. Entropy regularization. In: Chapelle O, Scholkopf B, Zien A, editors. Semi-Supervised Learning. Cambridge (MA): The MIT Press; 2006. p. 151–68.
23. Zappia L, Phipson B, Oshlack A. Splatter: simulation of single-cell RNA sequencing data. Genome Biol 2017;18:174.
24. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, De Boer CG, Jenkins RW, et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Cell 2018;175:998–1013.e20.
25. Bill R, Wirapati P, Messemaker M, Roh W, Zitti B, Duval F, et al. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. Science 2023;381:515–24.
26. Lauss M, Donia M, Svane IM, Jönsson G. B cells and tertiary lymphoid structures: friends or foes in cancer immunotherapy? Clin Cancer Res 2022;28:1751–8.
27. Zhang Y, Yu B, Ming W, Zhou X, Wang J, Chen D. SpaTopic: a statistical learning framework for exploring tumor spatial architecture from spatially resolved transcriptomic data. Sci Adv 2024;10:eadp4942.
28. Thrane K, Eriksson H, Maaskola J, Hansson J, Lundeberg J. Spatially resolved transcriptomics enables dissection of genetic heterogeneity in stage III cutaneous malignant melanoma. Cancer Res 2018;78:5970–9.
29. Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. Trends Immunol 2002;23:549–55.
30. Cheng S, Li Z, Gao R, Xing B, Gao Y, Yang Y, et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell 2021;184:792–809.e23.
31. Zhang T, Zhao F, Hu Y, Wei J, Cui F, Lin Y, et al. Environmental monobutyl phthalate exposure promotes liver cancer via reprogrammed cholesterol metabolism and activation of the IRE1α-XBP1s pathway. Oncogene 2024;43:2355–70.
32. Yang Z, Huo Y, Zhou S, Guo J, Ma X, Li T, et al. Cancer cell-intrinsic XBP1 drives immunosuppressive reprogramming of intratumoral myeloid cells by promoting cholesterol production. Cell Metab 2022;34:2018–35.e8.

33. Barras D, Ghisoni E, Chiffelle J, Orcurto A, Dagher J, Fahr N, et al. Response to tumor-infiltrating lymphocyte adoptive therapy is associated with preexisting CD8+ T-myeloid cell networks in melanoma. Sci Immunol 2024;9:eadg7995.
34. Chen JH, Nieman LT, Spurrell M, Jorgji V, Elmelech L, Richieri P, et al. Human lung cancer harbors spatially organized stem-immunity hubs associated with response to immunotherapy. Nat Immunol 2024;25:644–58.
35. Chu Y, Dai E, Li Y, Han G, Pei G, Ingram DR, et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. Nat Med 2023;29:1550–62.
36. Gide TN, Quek C, Menzies AM, Tasker AT, Shang P, Holst J, et al. Distinct immune cell populations define response to anti-PD-1 monotherapy and anti-PD-1/anti-CTLA-4 combined therapy. Cancer Cell 2019;35:238–55.e6.
37. Pauken KE, Shahid O, Lagattuta KA, Mahuron KM, Luber JM, Lowe MM, et al. Single-cell analyses identify circulating anti-tumor CD8 T cells and markers for their enrichment. J Exp Med 2021;218:e20200920.
38. Li N, Kang Y, Wang L, Huff S, Tang R, Hui H, et al. ALKBH5 regulates anti-PD-1 therapy response by modulating lactate and suppressive immune cell accumulation in tumor microenvironment. Proc Natl Acad Sci U S A 2020;117:20159–70.
39. Lu BY, Lucca LE, Lewis W, Wang J, Nogueira CV, Heer S, et al. Circulating tumor-reactive KIR+CD8+ T cells suppress anti-tumor immunity in patients with melanoma. Nat Immunol 2025;26:82–91.
40. Xue R, Zhang Q, Cao Q, Kong R, Xiang X, Liu H, et al. Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. Nature 2022;612:141–7.
41. Wang Y, Wang Q, Tao S, Li H, Zhang X, Xia Y, et al. Identification of SPP1+ macrophages in promoting cancer stemness via vitronectin and CCL15 signals crosstalk in liver cancer. Cancer Lett 2024;604:217199.
42. Fan G, Xie T, Li L, Tang L, Han X, Shi Y. Single-cell and spatial analyses revealed the co-location of cancer stem cells and SPP1+ macrophage in hypoxic region that determines the poor prognosis in hepatocellular carcinoma. NPJ Precis Oncol 2024;8:75.
43. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1:417–25.
44. Barcena-Varela M, Lujambio A. The endless sources of hepatocellular carcinoma heterogeneity. Cancers (Basel) 2021;13:2621.
45. Yang Y, Chen X, Pan J, Ning H, Zhang Y, Bo Y, et al. Pan-cancer single-cell dissection reveals phenotypically distinct B cell subtypes. Cell 2024;187:4790–811.e22.
46. Villéga F, Fernandes A, Jézéquel J, Uyttersprot F, Benac N, Zenagui S, et al. Ketamine alleviates NMDA receptor hypofunction through synaptic trapping. Neuron 2024;112:3311–28.e9.
47. Breslow NE. Analysis of survival data under the proportional hazards model. Int Stat Rev 1975;43:45–57.
48. Cox DR. Regression models and life-tables. J R Stat Soc Ser B (Methodol) 1972;34:187–202.
49. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184:3573–87.e29.
50. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15.
51. Balakrishnan V. All about the Dirac delta function (?) Resonance 2003;8:48–58.
52. Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, et al. beta-VAE: learning basic visual concepts with a constrained variational framework. 2017 [cited 2024 Jan 12]. Available from: https://openreview.net/forum?id=Sy2fzU9gl.
53. Csiszár I. I-divergence geometry of probability distributions and minimization problems. Ann Probab 1975;3:146–58.
54. Kingma DP, Salimans T, Jozefowicz R, Chen X, Sutskever I, Welling M. Improved variational inference with inverse autoregressive flow. In: Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, editors. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona (Spain): Curran Associates Inc.; 2016. p. 4743–51.
55. Brown PJ, Griffin JE. Inference with normal-gamma prior distributions in regression problems. Bayesian Anal 2010;5:171–88.
56. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics Intell Lab Syst 1987;2:37–52.
57. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al, editors. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach (CA): Curran Associates Inc.; 2017. p. 6000–10.
58. Yu T, Kumar S, Gupta A, Levine S, Hausman K, Finn C. Gradient surgery for multi-task learning. Adv Neural Inf Process Syst 2020;33:5824–36.
59. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87.
60. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–81.
61. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 1966;50:163–70.
62. Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag Process 2015;5:1.
63. Saikia R, Barman MP. A review on accelerated failure time models. Int J Stat Syst 2017;12:311–22.
64. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401:788–91.
65. Pölsterl S. Scikit-survival: a library for time-to-event analysis built on top of scikit-learn. J Machine Learn Res 2020;21:1–6.
66. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 2019;9:5233.
67. Zhao F, Zhang T, Wei J, Chen L, Liu Z, Jin Y, et al. Integrated single-cell transcriptomic analyses identify a novel lineage plasticity-related cancer cell type involved in prostate cancer progression. EBioMedicine 2024;109:105398.
68. Lefebvre C, Rajbhandari P, Alvarez MJ, Bandaru P, Lim WK, Sato M, et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol Syst Biol 2010;6:377.
69. Liu F, Zhang T, Yang Y, Wang K, Wei J, Shi JH, et al. Integrated analysis of single-cell and bulk transcriptomics reveals cellular subtypes and molecular features associated with osteosarcoma prognosis. BMC Cancer 2025;25:280.
70. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell 2016;165:35–44.
71. Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. Nature 2017;551:517–20.
72. Wang YP, Yu GR, Lee MJ, Lee SY, Chu IS, Leem SH, et al. Lipocalin-2 negatively modulates the epithelial-to-mesenchymal transition in hepatocellular carcinoma through the epidermal growth factor (TGF-beta1)/Lcn2/Twist1 pathway. Hepatology 2013;58:1349–61.
73. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res 2010;70:10202–12.
74. Cancer Genome Atlas Research Network. Electronic address wbe, cancer genome atlas research N. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 2017;169:1327–41.e23.
75. Kudo A, Mogushi K, Takayama T, Matsumura S, Ban D, Irie T, et al. Mitochondrial metabolism in the noncancerous liver determine the occurrence of hepatocellular carcinoma: a prospective study. J Gastroenterol 2014;49:502–10.
76. Ning J, Ye Y, Shen H, Zhang R, Li H, Song T, et al. Macrophage-coated tumor cluster aggravates hepatoma invasion and immunotherapy resistance via generating local immune deprivation. Cell Rep Med 2024;5:101505.