

Systems biology

scSurv: a deep generative model for single-cell survival analysis

Chikara Mizukoshi^{1,2,3,*} , Yasuhiro Kojima^{4,*}, Shuto Hayashi¹ , Ko Abe¹, Daisuke Kasugai^{5,6}, Teppei Shimamura^{1,2,*} 

¹Department of Computational and Systems Biology, Division of Biological Data Science, Medical Research Laboratory, Institute for Integrated Research, Institute of Science Tokyo, Tokyo 113-8510, Japan

²Division of Systems Biology, Graduate School of Medicine, Nagoya University, Nagoya, Aichi 466-8550, Japan

³Nagoya University Hospital, Nagoya, Aichi 466-8560, Japan

⁴Laboratory of Computational Life Science, National Cancer Center Research Institute, Tokyo, 104-0045, Japan

⁵Department of Emergency and Critical Care Medicine, Graduate School of Medicine, Nagoya University, Nagoya, Aichi 466-8550, Japan

⁶Institute of Nano-Life-Systems, Institutes of Innovation for Future Society, Nagoya University, Nagoya, Aichi 464-8601, Japan

*Corresponding authors. Chikara Mizukoshi, Department of Computational and Systems Biology, Division of Biological Data Science, Medical Research Laboratory, Institute for Integrated Research, Institute of Science Tokyo, Tokyo 113-8510, Japan; Division of Systems Biology, Graduate School of Medicine, Nagoya University, Nagoya, Aichi 466-8550, Japan; Nagoya University Hospital, Nagoya, Aichi 466-8560, Japan. E-mail: mizucosc@is.tokyo.ac.jp; komizucosc@gmail.com; Yasuhiro Kojima, Laboratory of Computational Life Science, National Cancer Center Research Institute, Tokyo, 104-0045, Japan. E-mail: yakojim@ncc.go.jp; Teppei Shimamura, Department of Computational and Systems Biology, Division of Biological Data Science, Medical Research Laboratory, Institute for Integrated Research, Institute of Science Tokyo, Tokyo 113-8510, Japan; Division of Systems Biology, Graduate School of Medicine, Nagoya University, Nagoya, Aichi 466-8550, Japan. E-mail: shimamura.csb@tmd.ac.jp; yura.csb@tmd.ac.jp

Associate Editor: Pier Luigi Martelli

Abstract

Motivation: Single-cell omics analysis has unveiled the heterogeneity of various cell types within tumors. However, no methodology currently reveals how this heterogeneity influences cancer patient survival at single-cell resolution. Here, we introduce scSurv, combining a Cox proportional hazards model with a deep generative model of single-cell transcriptome, to estimate individual cellular contributions to clinical outcomes.

Results: The accuracy of scSurv was validated using both simulated and real datasets. This method identifies cells associated with favorable or adverse prognoses and extracts genes correlated with their contribution levels. In melanoma, scSurv reproduces known prognostic macrophage classifications and facilitates hazard mapping through spatial transcriptomics in renal cell carcinoma. We also identified genes consistently associated with prognosis across multiple cancers and demonstrated the applicability of this method to infectious diseases. scSurv is a novel framework for quantifying the heterogeneity of individual cellular effects on clinical outcomes.

Availability: The implementation of scSurv is available on GitHub (<https://github.com/3254c/scSurv>) and Zenodo (<https://doi.org/10.5281/zenodo.17793054>).

Introduction

Survival analysis is a statistical method widely used across various fields to model the time until a specific event occurs, accounting for the influence of multiple covariates. These models typically consist of two primary components: the baseline hazard function, which represents the underlying hazard when all covariates are absent, and the effect parameters, which quantify how explanatory covariates influence the hazard function. Among them, the Cox proportional hazards model is the most widely used. This semi-parametric model assumes covariates have a multiplicative effect on the hazard function and allows the estimation of effect parameters without specifying the form of the baseline hazard function.

The Cox proportional hazards model has significantly enhanced the analysis of associations between molecular profiles-obtained through next-generation sequencing (NGS)

and mass spectrometry-and survival times, enabling the identification of prognostic factors and the development of predictive models (Gentles *et al.* 2015, Zhang *et al.* 2016, Thorsson *et al.* 2018). However, traditional approaches based on population averages are limited, particularly when addressing the cellular heterogeneity inherent in diseases and the specific role of various tissues and cell types in pathology. The diversity of cells and intercellular interactions within tumors and the tumor microenvironment influences disease progression and treatment responses (Chen and Mellman 2017, Dagogo-Jack and Shaw 2018, Rad *et al.* 2020), underscoring the need for more precise analyses at the cellular level.

Recent advances in single-cell sequencing technologies have enabled the high-resolution characterization of gene expression patterns and individual cell states, offering novel insights into disease mechanisms and identifying therapeutic

Received: 9 April 2025; Revised: 29 October 2025; Accepted: 2 December 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

targets (Regev *et al.* 2017, Papalexi and Satija 2018, Svensson *et al.* 2018). Nevertheless, large-scale cohorts integrating single-cell data with clinical information remain limited due to technical and economic constraints. In contrast, bulk RNA sequencing data linked to clinical outcomes are widely available. Despite their lower resolution, these data are valuable for analyzing relationships between molecular profiles and clinical outcomes.

Recent computational advancements have enabled the estimation of cell type proportions from bulk RNA sequencing data (Newman *et al.* 2019, Wang *et al.* 2019b, Chu *et al.* 2022). However, deconvolution models based on cell types inherently restrict analyses to this level, failing to capture variations in individual cell states or functions. These limitations hinder our ability to elucidate detailed biological mechanisms, including subtle shifts in cell states or associations between specific subtypes and survival outcomes. Consequently, critical biological insights may be overlooked, particularly in complex diseases like cancer, which exhibit pronounced intratumoral heterogeneity.

We addressed these limitations by developing scSurv, a novel method for single-cell survival analysis that extends the Cox proportional hazards model to incorporate cellular heterogeneity. Using single-cell RNA sequencing data as a reference, scSurv deconvolves bulk RNA-seq data to infer single-cell proportions and quantifies their individual contributions to patient outcomes. scSurv systematically identifies cells contributing to disease risk and their specific gene signatures by analyzing the association between survival time and cell abundance.

Simulations demonstrated that scSurv accurately predicts patient prognosis using single-cell proportions derived from deconvolution. Applying this method to The Cancer Genome Atlas (TCGA) data, we found that it could predict survival time of patients excluded from training across multiple cancers. scSurv identified specific cells influencing patient prognosis and the genes associated with these outcomes in melanoma. It also successfully reproduced known macrophage classifications affecting patient prognosis. Spatial transcriptomic analysis of renal cell carcinoma enabled tissue-wide hazard mapping and identified distinct prognosis-associated spatial regions. Furthermore, we identified genes contributing to prognosis across multiple cancers. Finally, applying this method to infectious diseases demonstrated its utility beyond cancer and survival outcomes. This novel approach elucidates the contributions of individual cells to clinical outcomes and offers new perspectives in clinical analyses.

Materials and methods

Concept of scSurv

We developed a novel method called scSurv, a deep generative model for single-cell survival analysis, to facilitate survival analysis at the single-cell level and to uncover biological mechanisms in diseases. scSurv involves the following steps: (1) This method uses single-cell RNA sequencing (scRNA-seq) data as a reference, following our previous framework (Kojima *et al.* 2024). Bulk RNA sequencing (bulk RNA-seq) data are deconvoluted at the cellular level using latent cell states obtained from a variational autoencoder (VAE). (2) scSurv estimates the hazard function using a Cox proportional hazards model, extended by combining the estimated proportion of each single cell within the bulk samples and the regression coefficients obtained from the latent cell state. These regression coefficients

are interpreted as the contributions of individual cells to clinical outcomes. This model enables the evaluation of hazard contributions at the single-cell level and enhances their consistency among cells with similar cell states.

The trained model offers three main analyses: (1) Quantification of individual cells' contributions to clinical outcomes. (2) Identification of prognosis-associated gene sets. (3) Mapping of spatial hazard distributions using spatial transcriptome data.

Through these applications, scSurv provides a comprehensive and interpretable framework that reveals heterogeneity in the clinical significance of cells. This framework enables the identification of novel cell populations and genes involved in the prognosis. This method is available as an open-source Python package on GitHub (<https://github.com/3254c/scSurv>).

scSurv framework

The scSurv framework consists of three main steps: learning latent states through VAE, deconvolution of bulk data, and estimation of regression coefficients in hazard functions for each cell using the extended Cox proportional hazards model. These steps are performed sequentially. The parameters learned in each step are then frozen and are not updated in the subsequent steps. The first two steps are similar to those in our previous study, DeepCOLOR (Kojima *et al.* 2024). scSurv employs a conditional VAE framework (Kingma *et al.* 2014) similar to scVI (Lopez *et al.* 2018) to handle scRNA-seq data collected from multiple patients as input. For all training steps, 85% of cells were randomly assigned to the training set, 10% to the validation set, and 5% to the test set. This same split was used for the VAE, deconvolution, and extended Cox steps. For the extended Cox step, bulk RNA-seq data were split into 60% training, 20% validation, and 20% test sets. The validation set is used for early stopping in order to suppress overfitting. The test set is used to compute reconstruction accuracy and the c-index with the trained model and to provide the final evaluation of model accuracy. The VAE compresses raw gene expression into low-dimensional latent cell representations, which can be treated as summaries of essential cellular information. scSurv utilizes the loss function of the Cox proportional hazards model to learn the contribution to the hazard function using the proportion of each single cell in each bulk sample as covariates. The proportions of the cells and their contributions to the hazard function learned in this process depends on the latent states of the cells. Consequently, cells with similar latent states are estimated to have similar contributions. This approach enables the model to effectively learn the contributions of 10,000 cells. A summary of the architecture, optimization settings, and hyperparameters is available in the [supplementary materials \(Table S1\)](#). These materials also include a detailed description of the model training process.

Derivation of the stochastic latent representation of the single-cell transcriptome

We define a probabilistic model for raw counts of single-cell transcriptomes. Let $\mathbf{z}_c \in \mathbb{R}^M$ represent the low-dimensional latent cell state, where M is the dimension of the latent space and the subscript c denotes each cell. We assume a Gaussian prior distribution for $\mathbf{z}_c$:

$$p(\mathbf{z}_c) = \mathcal{N}(\mathbf{z}_c | 0, I)$$

Let G be the number of genes and $\mathbf{x}_c \in \mathbb{R}^G$ be the raw counts of the single-cell transcriptome. We assume that given $\mathbf{z}_c$, the counts follow a Poisson distribution:

$$p(x_{cg} | \mathbf{z}_c, \mathbf{p}_c) = \text{Poisson}(x_{cg} | \hat{x}_{cg})$$

$$\hat{x}_c = l_c f_{\theta}(\mathbf{z}_c, \mathbf{p}_c)$$

Here, $l_c \in \mathbb{R}$ is the mean of all genes for each cell, $\mathbf{p}_c \in [0, 1]^P$ represents the batch information (patient information in this context), P is the total number of batches, and $p_{c,k} = 1$ indicates that cell c is collected from patient k . The function $f_{\theta}(\mathbf{z}_c, \mathbf{p}_c) \in \mathbb{R}^G$ is the decoding neural network of the latent cell state.

To obtain the latent cell state $\mathbf{z}_c$, we use a VAE framework to represent the posterior distribution $q(\mathbf{z}_c | \mathbf{x}_c, \mathbf{p}_c)$. We assume that the posterior distribution follows a Gaussian distribution:

$$q(\mathbf{z}_c | \mathbf{x}_c, \mathbf{p}_c) = \mathcal{N}(\mathbf{z}_c | \mu_{\Phi}(\mathbf{x}_c, \mathbf{p}_c), \text{diag}(\sigma_{\Phi}(\mathbf{x}_c, \mathbf{p}_c)))$$

Where $\mu_{\Phi}(\mathbf{x}_c)$ and $\sigma_{\Phi}(\mathbf{x}_c)$ denote the encoding neural networks.

We maximize the following Evidence Lower Bound to optimize the parameters of the generative model and variational distribution:

$$\mathcal{L}(\theta, \phi) = \sum_{c=1}^C \left(\log p(\mathbf{x}_c | \mathbf{z}'_c, \mathbf{p}_c) - D_{KL}(q(\mathbf{z}_c | \mathbf{x}_c, \mathbf{p}_c) || p(\mathbf{z}_c)) \right)$$

Where $\mathbf{z}'_c$ is obtained from $q(\mathbf{z}_c | \mathbf{x}_c, \mathbf{p}_c)$ using the reparameterization trick.

Probabilistic model of bulk transcriptome data and spatial transcriptome data

We model the expression $\mathbf{e}_b \in \mathbb{R}^G$ of bulk b by following a negative binomial distribution:

$$p(e_{bg} | \mu_{bg}^{bulk}, \alpha_g^{bulk}) = \text{NegativeBinomial}(e_{bg} | \mu_{bg}^{bulk}, \alpha_g^{bulk})$$

$$\mu_{bg}^{bulk} = r_g^{bulk} \sum_{c=1}^C f_{\theta}(\mathbf{z}_c, \mathbf{p}_c)_g m_{\theta^{bulk}}(\mathbf{z}_c)_b + o_g^{bulk}$$

Where $\mu_{bg}^{bulk} \in \mathbb{R}$ is the mean parameter, $\alpha_g^{bulk} \in \mathbb{R}$ is the dispersion parameter, $r_g^{bulk} \in \mathbb{R}$ represents the capture rate of each gene in bulk RNA-seq compared to scRNA-seq, and $o_g^{bulk} \in \mathbb{R}$ is the shift parameter for each gene. Both r_g^{bulk} and o_g^{bulk} are constrained to be positive scalar parameters shared across all bulk samples, ensuring model identifiability. $m_{\theta^{bulk}}(\mathbf{z}_c) \in \mathbb{R}^B$ is a neural network that outputs the proportion of each cell in each bulk sample given the latent cell state as the input, satisfying:

$$\sum_{c=1}^C m_{\theta^{bulk}}(\mathbf{z}_c)_b = 1$$

We optimize the parameters $\theta^{bulk}, \mathbf{r}^{bulk}, \mathbf{o}^{bulk}, \alpha^{bulk}$ by maximizing the following log-likelihood:

$$\mathcal{L}(\theta^{bulk}, \mathbf{r}^{bulk}, \mathbf{o}^{bulk}, \alpha^{bulk}) = \sum_{b=1}^B \log p(\mathbf{e}_b | \mu^{bulk}, \alpha^{bulk})$$

Similarly, we model the expression $\mathbf{e}_s \in \mathbb{R}^G$ of Visium spot s following a negative binomial distribution:

$$p(e_{sg} | \mu_{sg}^{spot}, \alpha_g^{spot}) = \text{NegativeBinomial}(e_{sg} | \mu_{sg}^{spot}, \alpha_g^{spot})$$

$$\mu_{sg}^{spot} = r_g^{spot} \sum_{c=1}^C f_{\theta}(\mathbf{z}_c, \mathbf{p}_c)_g m_{\theta^{spot}}(\mathbf{z}_c)_s + o_g^{spot}$$

Where the parameters are defined analogously to the bulk model, with $m_{\theta^{spot}}(\mathbf{z}_c) \in \mathbb{R}^S$, where S denotes the number of spots, satisfying:

$$\sum_{c=1}^C m_{\theta^{spot}}(\mathbf{z}_c)_s = 1$$

We optimize the parameters $\theta^{spot}, \mathbf{r}^{spot}, \mathbf{o}^{spot}, \alpha^{spot}$ by maximizing the following log-likelihood:

$$\mathcal{L}(\theta^{spot}, \mathbf{r}^{spot}, \mathbf{o}^{spot}, \alpha^{spot}) = \sum_{s=1}^S \log p(\mathbf{e}_s | \mu^{spot}, \alpha^{spot})$$

Single-cell cox proportional hazards model

We define the hazard function for each bulk sample as:

$$h(t)_b = h_0(t) \exp(\eta_b)$$

$$\eta_b = \sum_{c=1}^C \beta_{\theta'}(\mathbf{z}_c) m_{\theta^{bulk}}(\mathbf{z}_c)_b$$

Where $\beta_{\theta'}(\mathbf{z}_c) \in \mathbb{R}^c$ is a neural network that outputs regression coefficients given latent cell states as input, and $h_0(t)$ is the baseline hazard function, which remains unspecified.

Following the Cox proportional hazards model, we optimize the parameters θ' of $\beta_{\theta'}(\mathbf{z}_c)$ by maximizing the partial log-likelihood. Using the Breslow method (Breslow 1974), we maximize:

$$\mathcal{L}_{breslow}(\theta') = \sum_{b: I_b=1} \left(\eta_b - \log \sum_{i: T_i \geq T_b} \exp(\eta_i) \right)$$

Where $I_b = 1$ indicates the occurrence of an event (death) for the patient of bulk b . T_b represents the time of death when $I_b = 1$, or the last contact time when $I_b = 0$.

Using the Efron method (Efron 1977), we maximize:

$$\mathcal{L}_{efron}(\theta') = \sum_{b: I_b=1} \left\{ \eta_b - \log \left(\sum_{i: T_i \geq T_b} \exp(\eta_i) - \frac{r_b - 1}{d_b} \sum_{l: T_l = T_b, E_l = 1} \exp(\eta_l) \right) \right\}$$

Where d_b is the total number of patients experiencing events simultaneously at T_b , and r_b is the 1-based index indicating the position of bulk b in tied data, with $r_b = 1$ when $d_b = 1$.

Results

Validation using simulated datasets

We validated the performance of scSurv using simulated datasets. Whereas the existing bulk RNA-seq deconvolution

methods only perform cluster-level resolution, scSurv performs deconvolution at the single-cell level. Through simulations, we assessed the accuracy of scSurv's deconvolution and the precision of estimating the contributions to the hazard function based on the inferred cell proportions. Our results showed that under realistic conditions, single-cell level deconvolution achieves higher accuracy and is more effective for hazard regression than cluster-level methods.

We first clustered scRNA-seq data and assigned cluster labels to generate the simulated data. We then split the data into two subsets: one used as a reference and the other used to create pseudo-bulk samples. We randomly determined the proportion of each cell cluster within the bulk samples and generated pseudo-bulk RNA-seq data by aggregating their expression. We assigned regression coefficients in the hazard function to each cluster and set the survival times for each pseudo-bulk sample based on the Cox proportional hazards model.

We evaluated scSurv's performance in deconvoluting bulk data and compared it with existing methods such as CIBERSORTx (Newman *et al.* 2019), MuSiC (Wang *et al.* 2019b), and BayesPrism (Chu *et al.* 2022), which perform cluster-level deconvolution of bulk RNA-seq using scRNA-seq data as a reference. However, these methods do not achieve single-cell resolution. In realistic settings, the boundaries of meaningful cell populations may not align with the provided cluster labels. Therefore, we evaluated two scenarios: one in which we provided the same cluster labels used to generate the pseudo-bulk data for these methods and the other in which we provided labels with different boundaries.

The cell-type proportions estimated by scSurv were positively correlated with the ground-truth values in the simulation (Fig. 1A). When the existing methods were given the same cluster labels used to generate pseudo-bulk RNA-seq data, their estimated cluster-level proportions were more accurate than those of scSurv. However, when different labels were provided, scSurv's deconvolution accuracy was higher than the existing methods. This finding indicates that cluster-level deconvolution methods are limited by performance variability depending on the cluster labels provided and the clustering process. In contrast, scSurv does not use cell type annotations, so its performance is unaffected by the introduced noise. scSurv's single-cell level deconvolution avoids the bias introduced by clustering and produces consistent results.

Additionally, when estimating the contribution of each cluster to the hazard function using the inferred cell proportions, scSurv's estimates were positively correlated with the ground-truth values (Fig. 1B). In both cases, where the same cluster labels were used as those in generating the pseudo-bulk RNA-seq and where different labels were used, scSurv's estimation accuracy for contributions was higher than that of the existing methods. We found that cluster-level deconvolution methods produce results that vary depending on the clusters provided and fail to accurately estimate contributions when the cluster labels used in the hazard estimation differed from those used to determine hazard values in the simulation. These results validate the accuracy of scSurv's estimations and support the advantage of performing single-cell level estimation of the contribution to the hazard function.

Evaluating generalization performance using TCGA datasets

We applied scSurv to real datasets from TCGA. We selected 12 cancer types that had bulk RNA-seq data from over 300

patients to ensure learning stability and reported the c-index and Integrated Brier Score (IBS) (Fig. 1C, Supplementary Fig. S1E–S1F). In the training data, scSurv successfully predicted the hazard function based on cell proportions (Supplementary Fig. S1D). In the test data, scSurv demonstrated generalization performance by achieving a 95% confidence interval for the concordance index above 0.5 in multiple cancers (Fig. 1C). However, in some cancer types, the 95% confidence interval for estimation accuracy crossed 0.5, indicating difficulty in prediction. In these cancers, the confidence intervals of the c-index for the hazard estimation also crossed 0.5, even when we directly used the bulk RNA-seq data for hazard estimation, suggesting that bulk RNA-seq data may lack sufficient information to accurately predict the hazard function. This observation is consistent with previous studies that reported challenges in predicting prognosis using TCGA bulk RNA-seq data for certain cancer types (Cheerla and Gevaert 2019, Huang *et al.* 2020). Moreover, for the cancers with high estimation accuracy using bulk RNA-seq data (the lower bound of 95% CI > 0.5), scSurv's estimation accuracy outperformed the existing methods, except for BLCA. For the cancers with low prediction accuracy (the lower bound of 95% CI < 0.5), we found that the 95% CI of the c-index crossed 0.5 for most of the methods, suggesting that meaningful estimation of the hazard and their comparison were difficult in these cancers. From these findings, scSurv achieved accurate estimations across several cancer types, indicating that single-cell wise estimation of the hazard contribution by scSurv leads to more accurate prognosis prediction from bulk transcriptome data.

Identifying cells and genes associated with melanoma prognosis

We applied scSurv to a melanoma cohort (TCGA-SKCM). Melanoma is the most lethal type of skin cancers worldwide. (Leonardi *et al.* 2018) Our analysis revealed that specific populations of cancer cells, fibroblasts, endothelial cells, and macrophages adversely affected survival outcomes (Fig. 2A and B). We permuted the estimated contributions across cells within each cell cluster and quantified the decrement in the c-index to determine which cell clusters are important for prognostic prediction (Fig. 2C). Here, we found that cancer cells and macrophages exhibited a larger decrement, indicating that the heterogeneity of hazard contributions within these cell types had larger effects on the prognosis than the other cell types. Although intratumoral and intertumoral heterogeneity of cancer cells are documented factors influencing patient outcomes (Grzywa *et al.* 2017), recent findings additionally highlight the critical role of tumor-associated macrophages in shaping the tumor microenvironment and affecting patient prognosis (DeNardo and Ruffell 2019). Accordingly, our analysis focused on macrophages.

We isolated macrophages and identified genes whose expression correlated with scSurv estimated contributions (Fig. 2D). Gene set enrichment analysis revealed that interferon gamma signaling pathways were associated with favorable prognosis (Fig. 2E). This finding is consistent with those of previous studies (Wang *et al.* 2017). Additionally, the SPP1 gene characterizes macrophage subsets that contribute to tumor promotion (Bill *et al.* 2023). In contrast, the TNFSF10 gene encodes TNF-related apoptosis-inducing ligand (TRAIL), which promotes antitumor activity by inducing tumor cell apoptosis (Eisinger *et al.* 2020; Gunalp *et al.* 2023). In scSurv's estimations, SPP1 expression positively

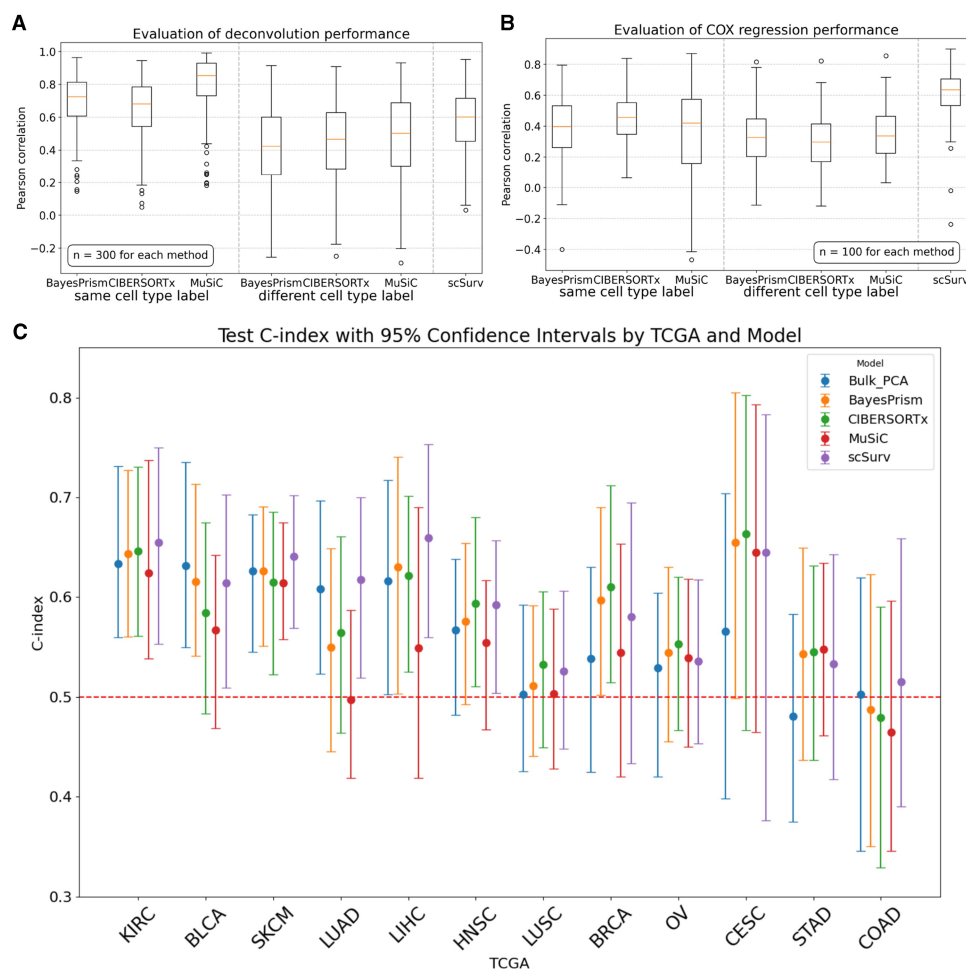


Figure 1. Evaluation of scSurv. **(A)** Deconvolution accuracy of scSurv, BayesPrism, CIBERSORTx, and MuSiC using 300 simulated pseudo-bulk samples. Pearson correlation between estimated proportions and ground truth values was calculated. BayesPrism, CIBERSORTx, and MuSiC were evaluated under two conditions: using same or different cluster labels between pseudo-bulk generation and deconvolution. **(B)** Accuracy of regression coefficient estimation using simulated data across methods. Pearson correlation between estimated and ground truth regression coefficients was calculated across 100 datasets, which were generated using the same 300 pseudo-bulk samples with different survival time settings by varying coefficient values. For BayesPrism, CIBERSORTx, and MuSiC, coefficients were estimated using the lifelines library based on deconvoluted cell type proportions under two conditions. **(C)** The performance of scSurv, the combination of the existing methods for bulk deconvolution with the Cox proportional hazards model, and the combination of bulk PCA with the hazard model were evaluated across 12 TCGA cancer types using the c-index. 95% confidence intervals were calculated from 100 iterations with different train/validation/test splits. scSurv achieved a test c-index > 0.5 in KIRC, BLCA, SKCM, LUAD, LIHC and HNSC. Lower performance in certain cancers suggests insufficient prognostic information in bulkRNA-seq data.

correlated with adverse prognosis, whereas TNFSF10 expression correlated negatively (Fig. 2F). These results indicate that scSurv can classify macrophages into prognostic subsets, aligning with previous studies.

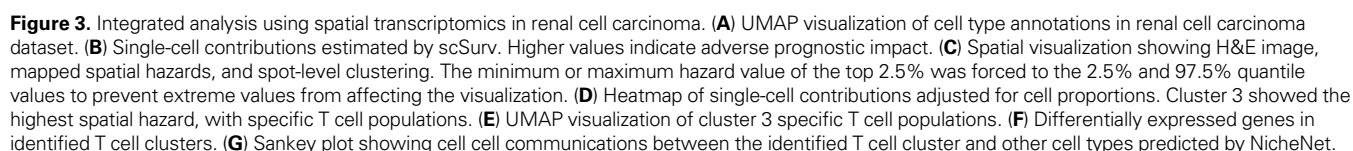
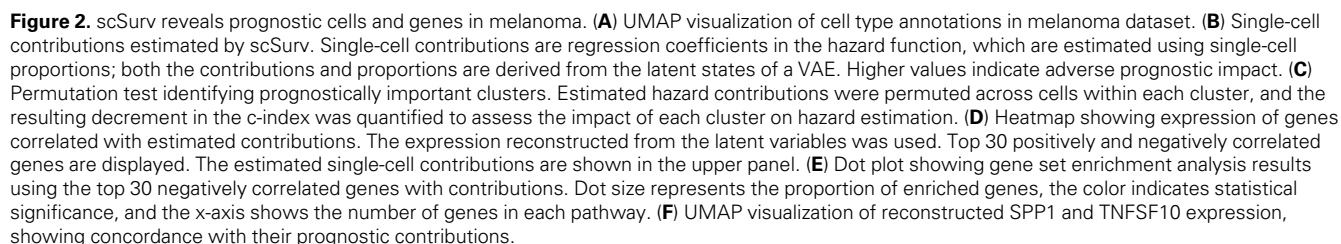
Our results confirm scSurv effectiveness in identifying prognostically relevant cells and genes. The consistency of these findings with existing knowledge highlights scSurv's utility in biological analyses.

Integration with spatial transcriptomics data in renal cell carcinoma

We expanded the scSurv analysis by integrating spatial transcriptomics data. First, we conducted the standard scSurv estimation to assign contributions to each cell using a renal cell carcinoma cohort (TCGA-KIRC) and scRNA-seq data (Li *et al.* 2022b) (Fig. 3A and B). Next, we estimated the cellular composition for each spot in the spatial transcriptome. We assigned a spatial hazard score to each spot based on single-

cell proportions and individual cell contributions (Fig. 3C). Regions with high spatial hazards were distributed between areas containing cancer cells and those containing normal cells (Supplementary Fig. S2A).

We clustered the spots based on their expressions to obtain spatial clusters, calculated the average hazard for each spatial cluster, and focused on clusters with particularly high hazards (Fig. 3D). Notably, we found that proliferative CD8-positive T cells specifically influenced the hazard function within this cluster (Fig. 3E, Supplementary Fig. S2B–S2D). This population of T cells is characterized by the expression of the Ki-67 gene (Fig. 3F) and is associated with poor prognosis in renal cell carcinoma (Blessin *et al.* 2021). Analysis of intercellular communication networks involving spatially co-localized cells in this population identified several prognostically relevant genes, including CCL4, CCL5, IFNG, ITGAM, and CXCL10 (Wang *et al.* 2019a, Zhang *et al.* 2021, Xu *et al.* 2022, Qu *et al.* 2022, Perelli *et al.* 2023) (Fig. 3G).



Integrating scSurv with spatial transcriptomics allowed us to perform hazard mapping across tissue sections, identify specific areas linked to patient prognosis, and examine the interactions between key cell populations.

Multiple cancer analysis

We performed a scSurv analysis of multiple cancer types. We selected six cancers where the lower bound of the 95% CI exceeded 0.5 in test patients and analyzed the cell populations and genes consistently linked to survival outcomes across these cancers. We first conducted permutation tests on the contribution scores of common cell types across the six cancers, including endothelial cells, fibroblasts, T cells, B cells, and myeloid cells, to determine their relative importance in the hazard function prediction (Fig. 4A). Among the analyzed cell types, myeloid cells had the strongest influence on prediction accuracy. We extracted myeloid cell populations from each cancer type and examined their correlations with the contributions of genes commonly expressed across all cancer types (Fig. 4B). We performed gene set enrichment analysis on the top and bottom genes ranked by mean correlation to elucidate the biological pathways consistently associated with prognosis in myeloid cells across different cancers. Genes associated with a good prognosis were enriched in antigen presentation-related terms (Fig. 4C), consistent with the known role of antigen presentation in antitumor immunity (Jhunjhunwala *et al.* 2021).

Next, we calculated the similarities between the cancer types based on their correlations (Fig. 4D). Notably, myeloid cells from hepatocellular carcinoma and renal cell carcinoma showed opposite correlation patterns. Analysis of differentially correlated genes (Fig. 4E) revealed enrichment of the Slit/Robo signaling pathways (Fig. 4F). These findings suggest that these pathways differentially affect survival outcomes in hepatocellular and renal carcinomas through mechanisms mediated by myeloid cells.

Application to other clinical outcomes in a COVID-19 cohort

We applied scSurv to hospitalized COVID-19 patients to demonstrate the applicability of our method to acute immune responses beyond cancer and to clinical outcomes beyond survival. Bulk RNA-sequencing data of peripheral blood mononuclear cells (PBMCs) and the corresponding clinical information were obtained from the IMPACC cohort (IMPACC Manuscript Writing Team and IMPACC Network Steering Committee 2021). Single-cell RNA sequencing data from PBMCs of 12 patients in Yoshida *et al.* (Yoshida *et al.* 2022) were used as a reference dataset (Fig. 5A). We applied scSurv to two distinct clinical outcomes: survival and discharge. scSurv provided robust predictions for both endpoints (Fig. 5B). Here, we found that monocytes exhibited a large contribution to survival hazard (Fig. 5C). Cellular contributions to survival hazard exhibited an inverse correlation with discharge hazard (Fig. 5D), reflecting the clinical

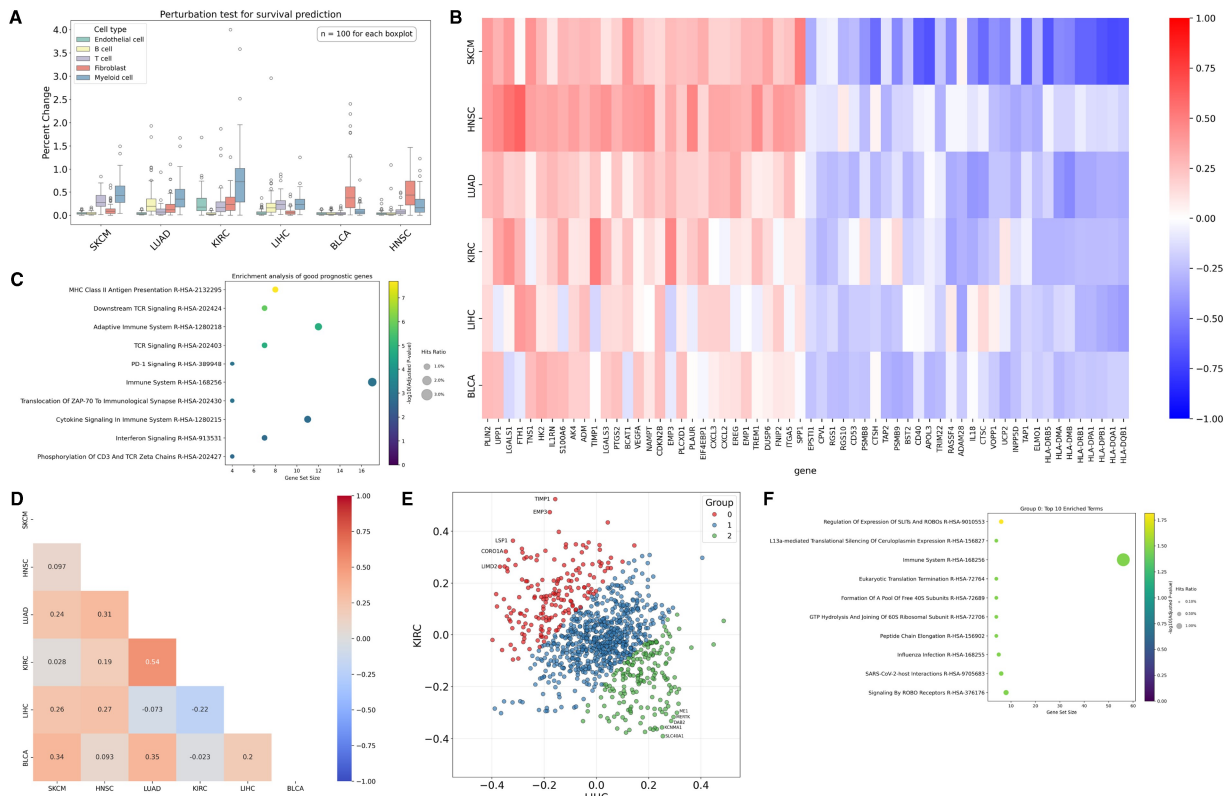


Figure 4. Pan-cancer analysis of stromal cells. (A) Permutation test identifying common prognostic cell types across six cancer types. (B) Heatmap showing Pearson correlations between myeloid cell gene expression and contributions across six cancer types. The heatmap illustrates the top 30 genes with the highest average correlation and the bottom 30 genes with the lowest average correlation across the cancer types. (C) Dot plot showing gene set enrichment analysis results using the top 30 negatively correlated genes with contributions. Dot size represents the proportion of enriched genes, the color indicates statistical significance, and the x-axis shows the number of genes in each pathway. (D) Cancer similarity based on gene correlations. (E) Scatter plot of correlations between gene expression and contributions in liver and kidney cancer myeloid cells. The genes were clustered into three groups by applying a Gaussian mixture model with three components to their projections onto the $y = -x$ line. (F) Gene set enrichment analysis results for differentially prognostic genes in liver and kidney cancer myeloid cells.

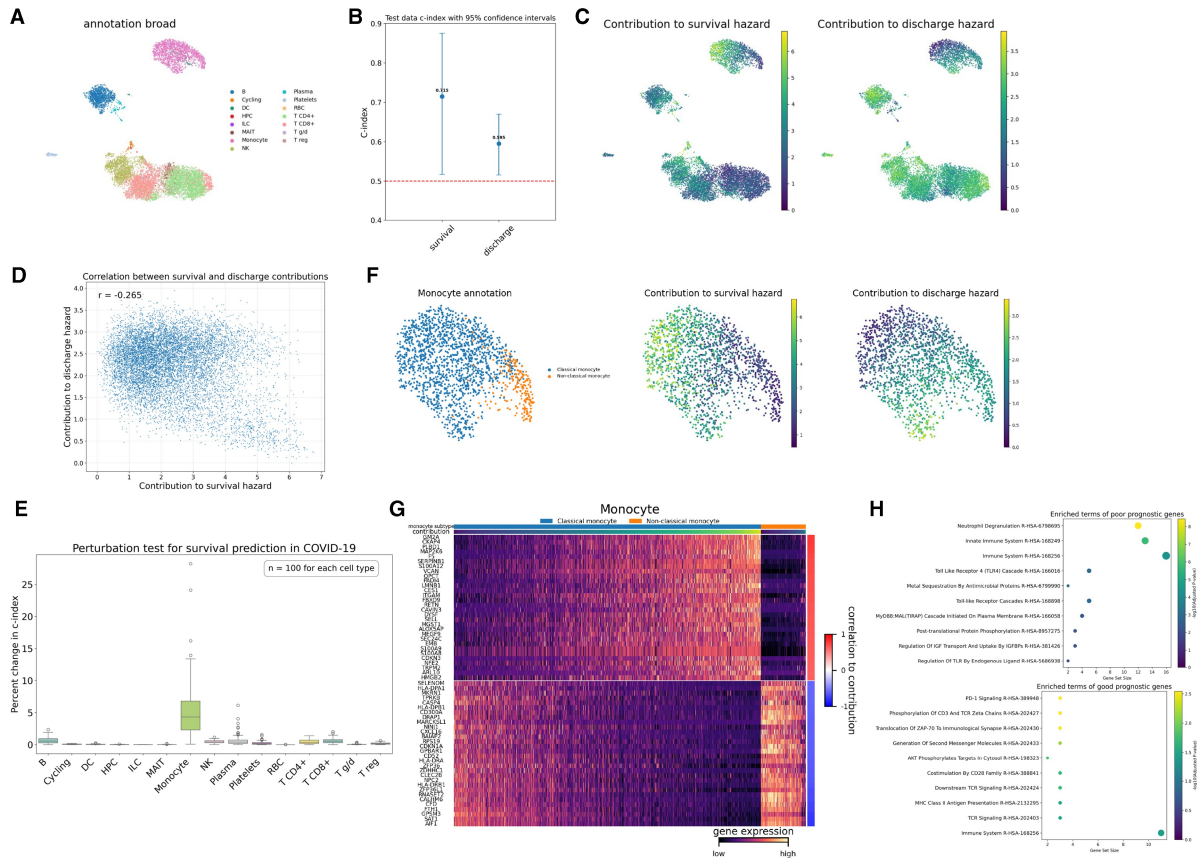


Figure 5. COVID-19 PBMC analysis. (A) UMAP visualization of cell type annotations in COVID-19 dataset. (B) 95% confidence intervals of scSurv c-indices for survival and discharge in test patients. (C) UMAP visualization of single-cell contributions to survival and discharge hazard. Higher values indicate adverse prognostic impact. (D) Scatter plot of single-cell contributions for both outcomes. Cellular contributions to survival hazard exhibited an inverse correlation with discharge hazard. (E) Permutation test identifying prognostically important clusters in COVID-19. (F) UMAP visualization of monocyte annotations and their hazard contributions to both outcomes. Classical monocytes show high contributions to survival hazard but low contributions to discharge hazard. (G) Heatmap showing reconstructed expression of genes correlated with estimated contributions to survival hazard. Top 30 positively and negatively correlated genes are displayed. The estimated single-cell contributions are shown in the upper panel. (H) Dot plot showing gene set enrichment analysis results using the top 30 positively and negatively correlated genes with contributions to survival hazard. Dot size represents the proportion of enriched genes, the color indicates statistical significance, and the x-axis shows the number of genes in each pathway.

relationship between shortened survival and extended hospitalization in severe disease. Permutation testing identified monocytes as the key cellular population in outcome prediction (Fig. 5E). These findings, including the large contribution of monocytes to the survival hazard and their importance identified through permutation testing, align with previous studies demonstrating the importance of monocytes in COVID-19 severity (Knoll *et al.* 2021, Vanderbeke *et al.* 2021). Based on these observations, we focused our subsequent analyses on this subset.

The distribution of contributions between classical and non-classical monocyte subtypes aligned with previous findings on the role of classical monocytes in COVID-19 severity (Vanderbeke *et al.* 2021) (Fig. 5F). The top 30 genes positively correlated with monocyte contributions included S100A12, S100A9, and S100A8, while the top 30 genes negatively correlated with monocyte contributions included HLA-DRA, HLA-DRB1, HLA-DPA1, and HLA-DPB1 (Fig. 5G). High expression of S100A proteins and low expression of HLA-DR proteins in monocytes have been identified as hallmarks of progressive disease (Unterman *et al.* 2022). These gene expression patterns were previously linked to disease severity in an analysis of scRNA-seq data from 66 PBMC samples (Knoll *et al.* 2024). While collecting scRNA-seq data from a large number of individuals is cost-prohibitive, scSurv

enables single-cell level analyses by integrating bulk RNA-seq data with a smaller scRNA-seq reference dataset. Our method provides a cost-effective solution for high-resolution profiling. Gene set enrichment analysis further revealed that genes positively correlated with monocyte contributions were enriched in neutrophil degranulation pathway and innate immune system pathway, aligning with reports linking activation of innate immune cells to poor COVID-19 prognosis (Fig. 5H) (Meizlish *et al.* 2021, Sievers *et al.* 2024). Conversely, genes negatively correlated with monocyte contributions were enriched in pathways related to TCR signaling and antigen presentation, supporting findings that impaired adaptive immune responses and antigen presentation are associated with disease severity (Moderbacher *et al.* 2020, Notarbartolo *et al.* 2021). These findings highlight scSurv's broader applicability beyond cancer and its utility in analyzing time-to-event data beyond survival time.

Discussion

We present the first methodology to quantify individual cells' contributions to clinical outcomes. This method allows the investigation of cells linked to patient outcomes using existing cohorts such as TCGA, enabling novel insights into

cellular heterogeneity at unprecedented resolution. scSurv does not rely on cell clustering, and its estimations are unaffected by cell labels, providing robust and unbiased results. Additionally, by utilizing the framework of conditional variational autoencoders for batch effect removal, we can effectively integrate scRNA-seq data from multiple patients to use as a reference.

Beyond its biological insights and practical utility, scSurv holds significance as an extension of Cox proportional hazards modeling. While notable deep learning-based Cox models such as DeepSurv (Katzman *et al.* 2018) and Cox-nnet (Ching *et al.* 2018) have enabled more accurate survival predictions, their non-linear nature presents challenges for interpretability. Although methods combining deep learning with linear predictors, such as Cox-PASNet (Hao *et al.* 2019) and PAGE-Net (Hao *et al.* 2020), have been developed, these approaches are not designed to address gene expression and heterogeneity at the individual cell level. scSurv extends the Cox proportional hazards model by incorporating both the estimated proportions of single cells in bulk RNA-seq samples, both derived from latent cell states obtained through a VAE. By leveraging these latent variables, this extended model enables precise survival analysis at the single-cell level and effectively scales to analyze contributions across large numbers of cells.

However, we recognize that scSurv has certain limitations. Similar to existing deconvolution methods, our method cannot assign proportions or contributions to cells that are not included in the reference. Therefore, it is essential to select reference datasets that comprehensively represent the cell populations found in bulk RNA-seq data. Additionally, the method requires a sufficient number of patients (over 300) and cannot be applied to cancers with rare death events. Furthermore, in certain cancer types, bulk RNA-seq data may contain insufficient predictive information, making accurate hazard function estimation difficult. Addressing these constraints remains a priority for future methodological development.

Despite these limitations, scSurv is a novel method to quantitatively evaluate individual cells' effects on clinical outcomes, enable identification of clinically relevant cell populations and genes, and generate new insights through integration with spatial transcriptomics. This method introduces an innovative concept of cellular heterogeneity based on contributions to clinical outcomes.

In summary, scSurv represents a significant advancement in single-cell analysis methodology, bridging the gap between cellular heterogeneity and clinical outcomes. We anticipate that scSurv will become an essential tool for researchers investigating the cellular basis of disease progression and treatment response, ultimately contributing to the development of more effective, personalized therapeutic strategies.

Acknowledgments

We gratefully acknowledge the IMPACC study group (IMPACC Manuscript Writing Team and IMPACC Network Steering Committee 2021) for providing bulk RNA-seq data and associated clinical outcome data used in this study.

Author contributions

Chikara Mizukoshi(Data curation [Lead], Formal analysis [Lead], Investigation [Lead], Methodology [Lead], Resources [Lead], Software [Lead], Validation [Lead], Visualization [Lead], Writing—original draft [Lead], Writing—review & editing [Lead]), Yasuhiro Kojima(Conceptualization [Lead], Funding acquisition [Supporting], Methodology [Lead], Supervision [Lead], Writing—review & editing [Lead]), Shuto Hayashi(Methodology [Supporting], Writing—review & editing [Supporting]), Ko Abe(Methodology [Supporting], Writing—review & editing [Supporting]), and Daisuke Kasugai(Data curation [Supporting], Resources [Supporting], Writing—review & editing [Supporting]), Teppei Shimamura (Funding acquisition [Lead], Project administration [Lead], Supervision [Lead], Writing—review & editing [Lead])

Supplementary data

Supplementary data is available at *Bioinformatics* online.

Conflict of interest: None declared.

Funding

The Grant-in-Aid for Transformative Research Areas (platforms for Advanced Technologies and Research Resources) [grant no. 22H04925] and Grant-in-Aid for Transformative Research Areas (A) [grant no. 23H04938] were provided by the Japan Society for the Promotion of Science (JSPS). Additional support for T.S. was received from the Japan Agency for Medical Research and Development (AMED) through the Core Research for Evolutional Science and Technology [grant no. JP25gm2010002], the Project Promoting Support for Drug Discovery [grant no. JP25nk0101112], Brain/MINDS Health and Diseases [grant no. JP25wm0625519], the Interdisciplinary Cutting-edge Research [grant no. JP25wm0325068], the Moonshot R&D Program [grant no. JP25zf0127012], and the Advanced Genome Research and Bioinformatics Study to Facilitate Medical Innovation (GRIFIN) [grant no. JP25tm0424226]. Further funding for T.S. was provided by the Japan Science and Technology Agency (JST) under the Moonshot R&D program [grant no. JPMJMS2025]. Y.K. was supported by the Project for P-PROMOTE [grant no. 24ama221609h0001] from AMED, the National Cancer Center Research and Development Fund [grant no. 2024-A-6], and a JSPS Grant-in-Aid for Early-Career Scientists [grant no. 23K16991]. Further support came from the Medical Research Center Initiative for High Depth Omics and Multilayered Stress Diseases at the Institute of Science Tokyo. Supercomputing resources were provided by the Shirokane supercomputer at the Human Genome Center, the University of Tokyo, and the TSUBAME3.0 supercomputer at the Institute of Science Tokyo.

Data availability

The implementation of scSurv is available on GitHub (<https://github.com/3254c/scSurv>) and Zenodo (<https://doi.org/10.5281/zenodo.17793054>). The single-cell RNA sequencing datasets used in this study are available from the Gene Expression Omnibus (GEO) under accession numbers GSE129845 (kidney cancer)(Li *et al.* 2022b), GSE115978 (melanoma)(Jerby-Arnon *et al.* 2018), GSE129845 (bladder

cancer)(Yu *et al.* 2019), GSE131907 (lung adenocarcinoma)(Kim *et al.* 2020), GSE176078 (breast cancer)(Wu *et al.* 2021), GSE132465 and GSE144735 (colorectal cancer)(Lee *et al.* 2020), GSE164690 (head and neck cancer)(Kürten *et al.* 2021), GSE149614 (hepatocellular carcinoma)(Lu *et al.* 2022), and GSE183904 (gastric cancer)(Kumar *et al.* 2022); from ArrayExpress under accession numbers E-MTAB-11948 (cervical cancer)(Li *et al.* 2022a) and E-MTAB-8107 (lung squamous cell carcinoma)(Qian *et al.* 2020); from Cell x Gene Explorer (Collection ID: 4796c91c-9d8f-4692-be43-347b1727f9d8) for ovarian cancer(Vázquez-García *et al.* 2022); and from COVID-19 Cell Atlas (<https://covid19cellatlas.org>) for COVID-19 PBMCs. Bulk RNA sequencing data and clinical information from TCGA are available from the Genomic Data Commons (GDC) Data Portal (<https://portal.gdc.cancer.gov/>). The bulk RNA-seq and clinical outcome data from the IMPACC cohort used in this study are accessible via the ImmPort database (<https://www.immport.org/>) with a data access request.

References

- Bill R, Wirapati P, Messemaker M *et al.* CXCL9: SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. *Science* 2023;381:515–24.
- Blessin NC, Li W, Mandelkow T *et al.* Prognostic role of proliferating CD8+ cytotoxic T cells in human cancers. *Cell Oncol (Dordr)* 2021;44:793–803.
- Breslow N. Covariance analysis of censored survival data. *Biometrics* 1974;30:89–99.
- Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics* 2019; 35:i446–i454.
- Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* 2017;541:321–30.
- Ching T, Zhu X, Garmire LX. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput Biol* 2018;14:e1006076.
- Chu T, Wang Z, Pe'er D *et al.* Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Cancer* 2022;3:505–17.
- Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* 2018;15:81–94.
- DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. *Nat Rev Immunol* 2019;19:369–82.
- Efron B. The efficiency of Cox's likelihood function for censored data. *J Am Stat Assoc* 1977;72:557–65.
- Eisinger S, Sarhan D, Boura VF *et al.* Targeting a scavenger receptor on tumor-associated macrophages activates tumor cell killing by natural killer cells. *Proc Natl Acad Sci U S A* 2020;117:32005–16.
- Gentles AJ, Newman AM, Liu CL *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 2015;21:938–45.
- Grzywa TM, Paskal W, Włodarski PK. Intratumor and intertumor heterogeneity in melanoma. *Transl Oncol* 2017;10:956–75.
- Gunalp S, Helvacı DG, Oner A *et al.* TRAIL promotes the polarization of human macrophages toward a proinflammatory M1 phenotype and is associated with increased survival in cancer patients with high tumor macrophage content. *Front Immunol* 2023;14: 1209249. This article is part of the Research Topic "Regulation of the Phenotype and Function of Human Macrophages and Dendritic Cells by Exogenous Immunomodulators".
- Hao J, Kim Y, Mallavarapu T *et al.* Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data. *BMC Med Genomics* 2019;12:189.
- Hao J, Kosaraju SC, Tsaku NZ *et al.* PAGE-Net: interpretable and integrative deep learning for survival analysis using histopathological images and genomic data. *Pac Symp Biocomput* 2020;25:355–66.
- Huang Z, Johnson TS, Han Z *et al.* Deep learning-based cancer survival prognosis from RNA-seq data: approaches and evaluations. *BMC Med Genomics* 2020;13:41.
- IMPACC Manuscript Writing Team and IMPACC Network Steering Committee. Immunophenotyping assessment in a COVID-19 cohort (IMPACC): a prospective longitudinal study. *Sci Immunol* 2021;6:eabf3733.
- Jerby-Arnon L, Shah P, Cuoco MS *et al.* A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* 2018;175:984–97.e24.
- Jhunjhunwala S, Hammer C, Delamarre L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat Rev Cancer* 2021;21:298–312.
- Katzman JL, Shaham U, Cloninger A *et al.* DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Methodol* 2018;18:24.
- Kim N, Kim HK, Lee K *et al.* Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun* 2020;11:2285.
- Kingma DP, Rezende DJ, Mohamed S *et al.* Semi-Supervised Learning with Deep Generative Models. *Adv Neural Inf Process Syst.* 2014;27:3581–9.
- Knoll R, Helbig ET, Dahm K, *et al.* The life-saving benefit of dexamethasone in severe COVID-19 is linked to a reversal of monocyte dysregulation. *Cell* 2024;187:4318–35.e20.
- Knoll R, Schultze JL, Schulte-Schrepping J. Monocytes and macrophages in COVID-19. *Front Immunol* 2021;12:720109.
- Kojima Y, Mii S, Hayashi S *et al.* Single-cell colocalization analysis using a deep generative model. *Cell Syst* 2024;15:180–92.e7.
- Kumar V, Ramnarayanan K, Sundar R *et al.* Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. *Cancer Discov* 2022;12:670–91.
- Kürten CHL, Kulkarni A, Cillo AR *et al.* Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat Commun* 2021;12:7338.
- Lee H-O, Hong Y, Etlioglu HE *et al.* Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet* 2020;52:594–603.
- Leonardi GC, Falzone L, Salemi R *et al.* Cutaneous melanoma: from pathogenesis to therapy (review). *Int J Oncol* 2018;52:1071–80.
- Li C, Wu H, Guo L *et al.* Single-cell transcriptomics reveals cellular heterogeneity and molecular stratification of cervical cancer. *Commun Biol* 2022;5:1208.
- Li R, Ferdinand JR, Loudon KW *et al.* Mapping single-cell transcriptomes in the intra-tumoral and associated territories of kidney cancer. *Cancer Cell* 2022;40:1583–99.e10.
- Lopez R, Regier J, Cole MB *et al.* Deep generative modeling for single-cell transcriptomics. *Nat Methods* 2018;15:1053–8.
- Lu Y, Yang A, Quan C *et al.* A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. *Nat Commun* 2022;13:4594.
- Meizlish ML, Pine AB, Bishai JD *et al.* A neutrophil activation signature predicts critical illness and mortality in COVID-19. *Blood Adv* 2021;5:1164–77.
- Moderbacher CR, Ramirez SI, Dan JM *et al.* Antigen-Specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. *Cell* 2020;183:996–1012.e19.
- Newman AM, Steen CB, Liu CL *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;37:773–82.

- Notarbartolo S, Ranzani V, Bandera A *et al.* Integrated longitudinal immunophenotypic, transcriptional, and repertoire analyses delineate immune responses in patients with COVID-19. *Sci Immunol* 2021;6:eabg5021.
- Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol* 2018;18:35–45.
- Perelli L, Carbone F, Zhang L *et al.* Interferon signaling promotes tolerance to chromosomal instability during metastatic evolution in renal cancer. *Nat Cancer* 2023;4:984–1000.
- Qian J, Olbrecht S, Boeckx B *et al.* A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res* 2020;30:745–62.
- Qu G, Wang H, Yan H *et al.* Identification of CXCL10 as a prognostic biomarker for clear cell renal cell carcinoma. *Front Oncol* 2022;12:857619.
- Rad HS, Monkman J, Warkiani ME *et al.* Understanding the tumor microenvironment for effective immunotherapy. *Medical Research Reviews* 2020;41:1474–98.
- Regev A, Teichmann SA, Lander ES, Human Cell Atlas Meeting Participants *et al.* The human cell atlas. *Elife* 2017;6:e27041.
- Sievers BL, Cheng MTK, Csiba K *et al.* SARS-CoV-2 and innate immunity: the good, the bad, and the “goldilocks”. *Cell Mol Immunol* 2024;21:171–83. published 20 november 2023
- Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc* 2018;13:599–604.
- Thorsson V, Gibbs DL, Brown SD, Cancer Genome Atlas Research Network *et al.* The immune landscape of cancer. *Immunity* 2018;48:812–30.e14.
- Unterman A, Sumida TS, Nouri N, *et al.* Single-cell multi-omics reveals dyssynchrony of the innate and adaptive immune system in progressive COVID-19. *Nat Commun* 2022;13:440. Published: 21 January 2022, Article number:
- Vanderbeke L, Mol PV, Herck YV *et al.* Monocyte-driven atypical cytokine storm and aberrant neutrophil activation as key mediators of COVID-19 disease severity. *Nat Commun* 2021;12:4117.
- Vázquez-García I, Uhlitz F, Ceglia N *et al.* Ovarian cancer mutational processes drive site-specific immune evasion. *Nature* 2022;612:778–86.
- Wang A, Chen M, Wang H *et al.* Cell Adhesion-Related molecules play a key role in renal cancer progression by multinetwork analysis. *Biomed Res Int* 2019a;2019:2325765.
- Wang H, Zhang L, Yang L *et al.* Targeting macrophage anti-tumor activity to suppress melanoma progression. *Oncotarget* 2017;8:18486–96.
- Wang X, Park J, Susztak K *et al.* Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* 2019b;10:380.
- Wu SZ, Al-Eryani G, Roden DL *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* 2021;53:1334–47.
- Xu W, Wu Y, Liu W *et al.* Tumor-associated macrophage-derived chemokine CCL5 facilitates the progression and immunosuppressive tumor microenvironment of clear cell renal cell carcinoma. *Int J Biol Sci* 2022;18:4884–900.
- Yoshida M, Worlock KB, Huang N, NU SCRIPT Study Investigators *et al.* Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature* 2022;602:321–7.
- Yu Z, Liao J, Chen Y *et al.* Single-Cell transcriptomic map of the human and mouse bladders. *J Am Soc Nephrol* 2019;30:2159–76.
- Zhang H, Liu T, Zhang Z, *et al.* Integrated proteogenomic characterization of human High-Grade serous ovarian cancer. *Cell* 2016;166:755–65.
- Zhang L, Zhang M, Wang L *et al.* Identification of CCL4 as an Immune-Related prognostic biomarker associated with tumor proliferation and the tumor microenvironment in clear cell renal cell carcinoma. *Front Oncol* 2021;11:694664.