nature portfolio

Corresponding author(s): DAPR NATCANCER-RS12035

Last updated by author(s): Mar 24, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Acquisition of immunofluorescence staining images from tissue sections used microscope (Leica, Microsystem)
Data analysis	All analyses were performed using R v4.0.5 and Python v3.7.1. The source codes can be retrieved from https://github.com/wanglabtongji/TabulaTiME. The following packages were used: MAESTRO v1.3.1, Seurat v4.3.0.1, inferCNV v1.3.3, Monocle v 2.18.0, Cellchat v1.1.3, CopyKAT v1.0.5, nichenetr v1.0.0 , GSVA v1.38.2, ROGUE v1.0, STRIDE v0.01 , ClusterProfiler v 3.18.1, SELINA v0.1, NMF v0.26, ggplot2 v3.4.3, survival v 3.5.7, ggpubr v0.6.0, ggsci v3.0.0, pheatmap v1.0.12, SCINA v1.2.0, xCell v1.1, TIMER v2.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data availability: Previously published datasets are available under their respective accession codes (Supplementary Tables 1 and 9). The analysis datasets in

TabulaTIME can be explored and visualized at http://wanglab-compbio.cn/TabulaTIME/. The data that support the findings of this study are available from the corresponding author upon request. For the original image data, users can download it from https://zenodo.org/records/13363711.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Neither sex nor gender was considered in the study design, since the primary focus of this study was unrelated to sex or gender. The sex was determined based on self-report. Regarding the published dataset, the clinical information was gathered from the original paper and will not be detailed here. For the mIHC analysis, there were seven patients - three males (NSCLC-1, NSCLC-2, and HNSCC-2) and four females (HNSCC-1, HNSCC-3, CESC-1, and CESC-3).
Reporting on race, ethnicity, or other socially relevant groupings	All the patients associated with mIHC analysis belong to the Asian race.
Population characteristics	The age of the patients NSCLC-1, NSCLC-2, HNSCC-1, HNSCC-2, HNSCC-3, CESC-1, and CESC-2 were 73, 55, 51, 74, 49, 55, and 44 years old, respectively. The clinical stage of the seven mIHC-associated patients includes three at stage I (NSCLC-1, HNSCC-2, and CESC-2), two at stage II (NSCLC-2 and HNSCC-3), one at stage III (CESC-1), and one at stage IV (HNSCC-1).
Recruitment	All sequencing-associated datasets analyzed herein are from patients presented in previously published studies. Patients analyzed via mIHC were selected based on a confirmed diagnosis of NSCLC, HNSC, or CESC, with no prior treatment and scheduled for curative-intent surgical resection.
Ethics oversight	Informed consent was obtained in writing from all human participants prior to tissue collection. The consent forms included disclosure of potentially identifiable information. Human tissue specimens were collected from the Shanghai Pulmonary Hospital (for the NSCLC sections), the West China Hospital of Stomatology at Sichuan University (for the HNSC sections), and the Shanghai East Hospital (for the CESC sections),under the approval of local medical ethics.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In our research, we focus on analyzing the tumor microenvironment of solid tumors. We collected publicly available scRNA-seq datasets associated with solid tumors, filtering the samples to include only those with more than 90% malignant cells, as samples with insufficient non-malignant cells were excluded. The samples were defined based on the original studies. After conducting data quality control, we compiled an extensive scRNA-seq dataset comprising 4,483,367 cells sourced from 746 patient samples across 36 cancer types, which include normal tissues, precancerous lesions, primary tumors, and metastatic sites. Following rigorous quality control procedures, we retained a total of 4,456,392 high-quality cells for further analysis. For specific analyses, the sample size was counted the total number of samples that met the criteria, such as being treatment-naive or including particular cell lineages.The corresponding sample sizes are indicated in the figure legends. To investigate the spatial distribution of the cell types of interest, we collected publicly available spatial transcriptomics data. Given that our analysis focuses on pan-cancer datasets, we gathered high-quality spatial transcriptomics data from 62 samples across six cancer types. Additionally, to validate the presence of specific cell types, we obtained tumor tissues from seven additional patients across three cancer types for mIHC staining, thereby enhancing the applicability of our findings across different cancers. To assess the clinical relevance of the identified cell types, we included data from 8,743 patients across 23 cancer types.These cancer types were selected because TabulaTIME includes corresponding scRNA-seq data from the TCGA project in our study.
Data exclusions	Considering potential technological biases, we concentrated on the compilation of datasets from 10X Genomics. For the scRNA-seq data collected from published studies, we included only high-quality solid tumor datasets that contained over 1,000 cells. This rigorous selection process culminated in the integration of data from 103 studies, encompassing 36 distinct cancer types, 746 donors, and a total of 4,483,367 cells. Cells with a UMI count below 1,000, fewer than 500 expressed genes, and those exhibiting more than 15% mitochondrial gene expression were excluded from further analysis. Furthermore, to identify and eliminate doublets, we utilized Scrublet with its default parameter of 0.06 for each dataset, resulting in the exclusion of 26,975 cells and leaving 4,456,392 cells for subsequent examination. To delineate tissue structures, we incorporated spatial transcriptomics data from 62 patients across six cancer types. To ensure data quality, we filtered spots based on a minimum detection threshold of 250 genes and a maximum mitochondrial gene expression of 25%. Additionally, we excluded genes that had fewer than 10 read counts or were expressed in fewer than two spots. Moreover, to evaluate the clinical relevance of the identified cell types, we included data from 8,743 tumor patients representing 23 cancer types sourced from the TCGA project in our analysis.

Replication	To validate the co-existence of eFibro_CTHRC1 and Macro_SLPI across different cancer types, we analyzed tumor tissues from three HNSCC patients, two NSCLC patients, and two CESC patients. Detailed information regarding the replicates is provided in Supplementary Table 8. Upon staining the same cell type markers in seven patients, we observed co-localization of eFibro_CTHRC1 and Macro_SLPI in the HNSCC and NSCLC patients. In contrast, in the CESC patients, eFibro_CTHRC1 was more likely to co-localize with Macro_SPP1 rather than Macro_SLPI. This lack of confirmation in the CESC samples highlights the diversity of tumor microenvironments across distinct cancer types.
Randomization	Randomization is not relevant to this study. Given that the study focuses on patients with solid tumors, a targeted selection of data was essential, rendering random sampling inappropriate. As a result, the study is designed as observational. In observational studies, researchers do not manipulate variables, which makes randomization unnecessary.
Blinding	Blinding is not relevant to this study. In this study, we focus on identifying and dissecting pro-tumor associated cell types. Since the nature of the study is purely observational rather than experimental, blinding is typically less applicable, as there is no intervention being tested. Additionally, the effects of specific cell types on tumor progression can be measured objectively and quantified easily. Therefore, blinding may not significantly impact the results, as the risk of bias is minimized in this context.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Multiplex immunohistochemistry: antibodies anti-CTHRC1 Abcam Cat# ab85739, RRID:AB_10712489 1:400 anti-SLPI ThermoFisher Cat# PA5-82990, RRID:AB_2790146 1:500 anti-osteopontin Abcam Cat# ab214050, RRID: AB_2894860 1:1500 anti-CD68 Biolynx Cat# BX50031, RRID: AB_2936308 1:400 anti-cytokeratin pan Biolynx Cat# BX50143 1:300
Validation	The antibodies used in this study were tested by their respective manufacturers. Anti-CTHRC1: https://www.abcam.com/products/primary-antibodies/cthrc1-antibody-ab85739.html [application: IHC-P; species: human] Anti-SLPI: https://www.thermofisher.com/cn/zh/antibody/product/SLPI-Antibody-Polyclonal/PA5-82990 [application: IHC-P; species: human] Anti-osteopontin: https://www.abcam.com/products/primary-antibodies/osteopontin-antibody-epr21139-316-ab214050.html [application: IHC-P; species: human] Anti-CD68: https://www.biolynxtec.com/products/antibody/cd68.html [application: IHC-P; species: human] Anti-cytokeratin pan: https://www.biolynxtec.com/products/antibody/cytokeratin-pan.html [application: IHC-P; species: human]

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA