SOAR: a spatial transcriptomics analysis resource to model spatial variability and cell type interactions

Yiming Li[1,#], Saya Dennis[1,#], Meghan R. Hutch[1,#], Yawei Li[1], Mia S. Broad[1], Zexian Zeng[2,3,*] and Yuan Luo[1,*]

[1] Department of Preventive Medicine, Northwestern University, USA
[2] Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China
[3] Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China

* Corresponding authors

[#] Co-first authors

Email addresses:

zexianzeng@pku.edu.cn (Zexian Zeng)

yuan.luo@northwestern.edu (Yuan Luo)

ABSTRACT

Spatial transcriptomics is an emerging technology to profile spatially resolved gene expression. Its preserved capture locations allow researchers to investigate transcriptomes in the tissue context. Large volumes of spatial transcriptomics data under different study designs have been generated, but the lack of a public database with systematically collected and processed data makes data reuse challenging. We present Spatial transcriptOmics Analysis Resource (SOAR), a database with analysis capability and an extensive collection of spatial transcriptomics data. We systematically curated, reviewed, annotated, and pre-processed 132 datasets containing 1,633 samples across 22 tissue types from 6 species. SOAR provides interactive web interfaces for users to visualize spatial gene expression, evaluate gene spatial variability across cell types, and assess cell-cell interactions. Besides data access and download, SOAR can aid researchers in investigating whether a specific gene expression is associated with distinct spatial patterns or cell-cell communications. SOAR is publicly available at https://soar.fsm.northwestern.edu/.

INTRODUCTION

Recent technological advances in spatial transcriptomics have made it possible to measure the transcriptome while retaining the coordinates of capture locations in tissues. Spatially resolved transcriptomics allows researchers to study the association between gene expression and the spatial organization of capture locations [1]. This provides helpful insights into tissue functions and disease pathology, facilitating discoveries in cancer research [2-9], neuroscience [2,10-23], and developmental biology [24-28]. Spatial transcriptomics can also play an important role in generating insights in precision medicine which increasingly integrates Omics data with other modalities of health care data (e.g., clinical data) [29-33]. In recent years, various spatial transcriptomics technologies have been proposed and widely applied, including next-generation sequencing (NGS) approaches like 10x Visium, ST [2], Slide-seq [10,24], and DBiT-seq [26], as well as fluorescence in situ hybridization (FISH) methods like MERFISH [34,35], osmFISH [13], seqFISH [11,34,36], and seqFISH+ [37]. Despite the rapid accumulation of spatial transcriptomics publications and datasets, the vastly dissimilar formats of datasets from different techniques make data reuse challenging. A user-friendly and comprehensive public database for spatial transcriptomics with analysis capability will greatly facilitate data sharing, exploration and meta-analysis among the research community.

A limited number of public spatial transcriptomics resources is available, including SpatialDB [38], Museum of Spatial Transcriptomics [39], spatialLIBD [12], and STAR-FINDer [25], Amyotrophic Lateral Sclerosis Spinal Cord Atlas [40]. SpatialDB [38] is a manually curated database with 24 datasets that provides spatial transcriptomics data visualization and spatially variable gene identification results. Museum of Spatial Transcriptomics [39] is an annotated literature list with download links to public spatial transcriptomics datasets. The other resources [12,25,40] are data atlases regarding one tissue type and provides spatial expression visualisations of brain, fetal intestine, and spinal cord, respectively. To the best of our

knowledge, there is a lack of a large-scale, systematically curated spatial transcriptomics database that provides unified data access, exploration and analysis across different tissue types.

Here, we present SOAR (Spatial transcriptOmics Analysis Resource), an extensive and publicly accessible resource of spatial transcriptomics data. SOAR (https://soar.fsm.northwestern.edu/) is a comprehensive database hosting a total of 1,633 samples from 132 datasets, which were uniformly processed using a standardized workflow. Its data collection covers 22 different tissue types including numerous organs, developmental stages, and pathological conditions such as cancer. SOAR also provides interactive web interfaces for users to visualize spatial gene expression, explore gene spatial variability, and assess cell-cell interactions using an in-house developed, novel approach. SOAR will be continuously maintained in order to provide great utility to the biological, biomedical, and clinical research communities for harnessing the power of spatial transcriptomics data.

MATERIAL AND METHODS

Data collection

We queried the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) for human and mouse spatial transcriptomics datasets using the keywords "spatial+transcriptomics", "spatial+transcriptome", "spatial+RNA-seq", and "spatial+RNA+sequencing", and downloaded 353 datasets from unique GEO series (GSE) accessions. Additionally, we manually reviewed the papers in the Museum of Spatial Transcriptomics [39] and collected 73 publicly available datasets (19 FISH, 54 NGS). We also collected 114 datasets from other resources including Single Cell Portal (https://singlecell.broadinstitute.org/single_cell), 10x spatial gene expression demonstration datasets (https://support.10xgenomics.com/spatial-gene-expression/datasets), Spatial

Research Lab (https://www.spatialresearch.org/resources-published-datasets/), 10x spatial publication list (https://www.10xgenomics.com/resources/publications), Amyotrophic Lateral Sclerosis Spinal Cord Atlas [40], spatialLIBD [12], STAR-FINDer [25], and Brain Research through Advancing Innovative Neurotechnologies Initiative – Cell Census Network (https://biccn.org/data). Next, we removed the duplicative datasets, validated that the downloaded data used FISH or NGS-based spatial transcriptomics technology, and excluded the datasets missing spatial coordinates information. In total, we collected 132 datasets containing 1,633 spatial transcriptomics samples from eight different technologies (Figure 1A).

Data processing

We downloaded the count matrices and coordinate information for each dataset and applied a systematic data processing workflow (Figure 1A) to all the collected datasets. To account for the resolution and sequencing depth difference among spatial transcriptomics techniques, samples measured using different technologies were processed with different quality control (QC) protocols. For 10x Visium, ST and DBiT-seq datasets, we removed the capture locations with fewer than 500 unique molecular identifiers (UMIs), fewer than 500 genes, or $\geq$ 25% mitochondrial reads [41,42]. We further excluded the capture locations with a total UMI count (or a total number of genes) three standard deviations below the medium [43]. Finally, we filtered out the genes that are expressed in less than five capture locations. The QC pipeline for MERFISH, osmFISH, seqFISH, and seqFISH+ datasets was similar. Genes expressed in fewer than five capture locations were excluded [43], and capture locations with fewer than 500 UMIs or $\geq$ 25% mitochondrial reads were removed [41]. We performed QC on Slide-seq samples so that only the capture locations with total UMI counts greater than 100 and the genes with UMI count greater than 300 in all the capture locations are included [38]. After QC, we normalized and transformed the datasets with raw, unnormalized expression matrices

using SCTransform, a framework for the normalization and variance stabilization of molecular count data [44]. We next performed principal component analysis on normalized data and clustered the capture locations through a shared nearest neighbour approach. All data processing was conducted using Seurat V3 [43].

Cell type annotation

We performed cell type annotation on SOAR's spatial transcriptomics datasets using SingleR [45], a method capable of annotating the cells in test datasets based on their similarities to reference single-cell RNA sequencing (scRNA-seq) datasets with known cell types. To identify such reference datasets, we queried the GEO and curated an average of two well-annotated scRNA-seq datasets for each tissue type featured in SOAR. After QC, normalization, and transformation, these scRNA-seq datasets were used as references for annotating the cell types of spatial transcriptomic capture locations of the corresponding tissue type. In particular, for complex tissues such as tumour and brain, we adopted heuristics-guided approaches to improve the performance of cell type annotation.

For cancer datasets (185 samples in total), we first identified possible lymphocyte capture locations using known biomarkers (CD45, CD3D, CD3E, and CD3G) through differential expression analysis, identifying clusters with greater than 1.2 average log fold-changes for one or more lymphocyte genes and over 20% capture locations expressing at least one of the biomarkers. Next, we identified capture location clusters that differentially expressed CD4 or CD8A/CD8B genes in a similar manner. Among these clusters, we annotated those that were also found to be possible lymphocyte clusters in the first step as CD4 or CD8 based on the percentage of capture locations in the cluster expressing CD4 versus CD8A/CD8B genes. Next, we performed cell type annotation on all the remaining clusters using SingleR [45], with scRNA-seq datasets of the same cancer type curated in a previous study [46] as the reference.

The cell type of each cluster was then determined by the annotation results from different references, weighted by the number of cells in the respective reference dataset.

For brain datasets (866 samples in total), we followed the Common Cell Type Nomenclature (CCN) [47] and annotated their capture locations as glutamatergic, GABAergic, or non-neuronal. Two scRNA-seq datasets from the Allen Brain Map (https://portal.brain-map.org/atlases-and-data/rnaseq) were used as the references – the Human Multiple Cortical Areas SMART-seq dataset (for annotating human samples) and the Mouse Whole Cortex and Hippocampus dataset (for annotating mouse samples). Firstly, we identified marker gene sets for each cell type and each species by performing differential gene expression analysis on the corresponding reference scRNA-seq dataset using Seurat V3 [43]. Next, we used AUCell [48] to score the activity of glutamatergic, GABAergic, and non-neuronal gene sets at each capture location based on marker gene expressions. Capture location clusters in the sample can then be classified as neuronal or non-neuronal according to the sum of AUCell scores across capture locations. Finally, we used SingleR [45] to annotate the neuronal clusters as glutamatergic or GABAergic based on a filtered version of the reference dataset that only contained neuronal cells.

Data analysis

To facilitate the characterization of the functional architecture of complex tissues, we identified genes with spatial patterns of significant expression variation using SPARK-X [49]. Spatial variability analyses were conducted across the whole tissue and in different cell types, respectively.

Cells of different cell types may interact through cell-cell contact or long-distance signalling [50]. To study possible cell type interactions, we investigated whether the gene expression levels in a query cell type (CT_Q) are influenced by its spatial proximity to another cell type (CT_I). In order to evaluate neighbouring interactions, we performed Wilcoxon rank-

sum tests to test if genes are differentially expressed in CT_I adjacent and nonadjacent to CT_Q using the FindAllMarkers function in Seurat V3 [43]. Next, to assess the cell type interactions over distance, we regressed the expression of a gene G in CT_Q ($E_{CTQ,G}$) capture locations on their median Euclidean distance to CT_I capture locations ($Distance_{CTQ,CTI}$). Intuitively, the calculated regression coefficient (β) quantified the association between the gene expression in CT_Q capture locations and the distance between CT_Q and CT_I. Therefore, a positive β may reflect CT_I's inhibitory effect on the expression of G in CT_Q, whereas when $\beta < 0$, CT_I capture locations may play a promotional role.

$$E_{CT_Q,G} = \alpha + \beta \cdot Median(Distance_{CT_Q,CT_I})$$

All the p-values were adjusted for multiple testing using the false discovery rate (FDR) approach, and we assume statistical significance at an adjusted p-value of $q < 0.05$.

Website development

SOAR is a comprehensive and user-friendly database that aids the exploration and analysis of spatial transcriptomics datasets. The website was implemented using the R Shiny framework (R v4.0.5) on an Apache2 HTTP server and is compatible with smartphones and tablets. The website consists of five functional components, "Home", "Data Browser", "Explore Gene", "Download", and "Help". The "Home" module includes an overview of SOAR, and users may search for a gene of interest in this module. Users could browse SOAR's curated datasets using the "Data Browser" module to pinpoint their sample of interest and visualize spatial gene expressions. Upon searching for a gene on the homepage, users will land in the "Explore Gene" module, which enables users to evaluate the spatial variability of genes in different tissues and assess possible cell type interactions. All the results and visualizations from user-performed analyses are downloadable. In the "Download" module, users could download phenotypic metadata, standardized gene expression, and coordinate data of all the

samples in SOAR. The "Help" page documents the website and includes a tutorial with step-by-step instructions for using the database. SOAR is free and open to all users at https://soar.fsm.northwestern.edu/ and there is no login requirement.

RESULTS

Data summary

SOAR includes 1,633 spatial transcriptomics samples of six different species (human, mouse, chicken, pig, thale cress, and aspen) from 132 datasets utilizing eight different spatial transcriptomics technologies (Figures 1B and 1C). The human and mouse samples come from different organs (brain, heart, intestine, joints, kidney, liver, lymph node, muscle, prostate, skin, spinal cord, testis) and other specific tissues including body fat, cancer, and embryonic tissues (Figure 1C). All the curated gene expression data, coordinate data, metadata, and analysis results can be downloaded from the "Download" module in SOAR.

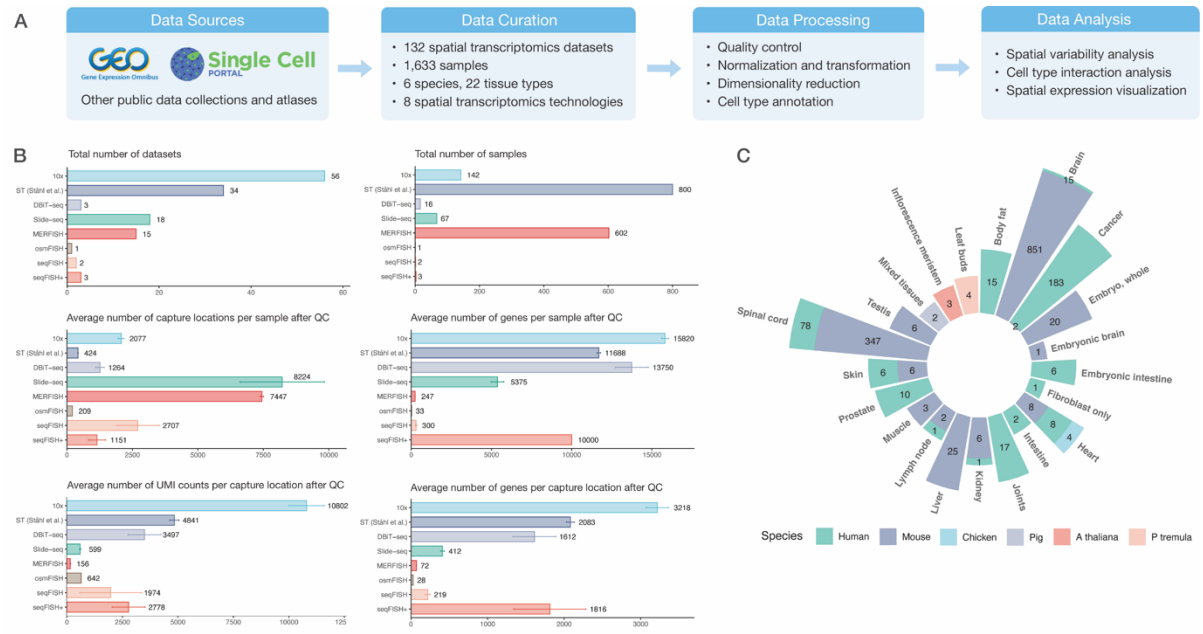

Figure 1. Overview of SOAR. (A) Datasets hosted on SOAR were curated from public domains. The samples were processed using a standardized workflow, including quality

control, normalization, transformation, dimensionality reduction, and cell type annotation. SOAR also provides interactive interfaces for users to evaluate the spatial variability of genes in a cell type of interest, assess cell-cell interactions, and visualize spatial gene expressions. The expression profiles and coordinate data for all samples can be downloaded from SOAR. **(B)** Statistics of data from different spatial transcriptomics technologies in SOAR. The 95% confidence intervals for the means are plotted as error bars. **(C)** In total, SOAR contains 1,633 spatial transcriptomics samples from 6 different species across 22 tissue types.

Data browser module

To aid user conducted analysis, we constructed a comprehensive data browser that is hosted on SOAR and contains the meta-data for all included spatial transcriptomics datasets. For each dataset, detailed information includes the hyperlink to the corresponding publication, the spatial transcriptomics technology used, and sample information including the number of samples, the species, organ, tissue, and the disease state of the sample. Furthermore, we document the average number of capture locations and genes per dataset, and in each sample after QC. Our data browser allows users to quickly select samples of interest to further explore and analyze via interactive figures and tables. All the generated figures and tables are easily downloadable to support personal and large-scale research projects.

Explore gene module

Spatial transcriptomics makes it possible to analyze a gene's spatial variability within different cell types. It also enables us to study cell type interactions by investigating whether a gene's expression appears to be promoted or inhibited when in proximity to another cell type. The gene search bar on the homepage of SOAR allows users to query the results of these analyses for a specific gene of interest. Upon searching for a gene, SOAR directs the user to the "Explore Gene" tab, which subsequently prompts the user to narrow down the list

of datasets by tissue type, species, and, for users conducting cancer studies, the type of cancer. SOAR allows users to perform three types of analyses – spatial variability, adjacency-based cell type interaction, and distance-based cell type interaction.

Spatial variability. The spatial patterns of gene expression variation may inform us about cell-cell interactions or the migration of cells to different tissue locations [51]. In the "Explore Gene" Module, the "Spatial variability" function allows users to evaluate the statistical significance of gene's spatial variability across the whole tissue or in specific cell types. When the user selects a gene, SOAR generates a heatmap to summarize the significance of the gene's spatial variability (Figure 2A). This heatmap allows users to explore whether the given gene has a distinct spatial expression pattern in the whole tissue or in a specific cell type. CXCL9 and CXCL13 are known to be associated with the degree and prognosis of cancer [52,53], and their spatial expressions were also found to have a significant pattern variation in previous studies [51,54]. As an illustration, we visualized the spatial variability of these genes across the whole tissue and in different cell types in breast cancer samples (Figure 2A). As shown in the figure, CXCL9 and CXCL13 often had significant spatial expression variations in malignant capture locations.

To improve the accessibility of this visualization function, SOAR also allows users to visualize the spatial expression patterns of a gene in the "Data Browser" module. For example, in a specific breast cancer sample, SOAR identified 6,757 spatially variable genes (FDR < 0.05), and they included cytokines like CXCL9 (adjusted p-value = 5.76×10^{-29}) and CXCL13 (adjusted p-value = 1.00×10^{-32}). These genes were found to be expressed in a visually distinct region (Figure 2B), along with the interleukin 2 receptor subunit gene IL2RB (adjusted p-value = 3.13×10^{-18}), indicating a potential acute local inflammatory response [55]. Taken together, these demonstrated that SOAR could provide useful insights for identifying biologically relevant biomarkers in tissue samples without histological annotation.

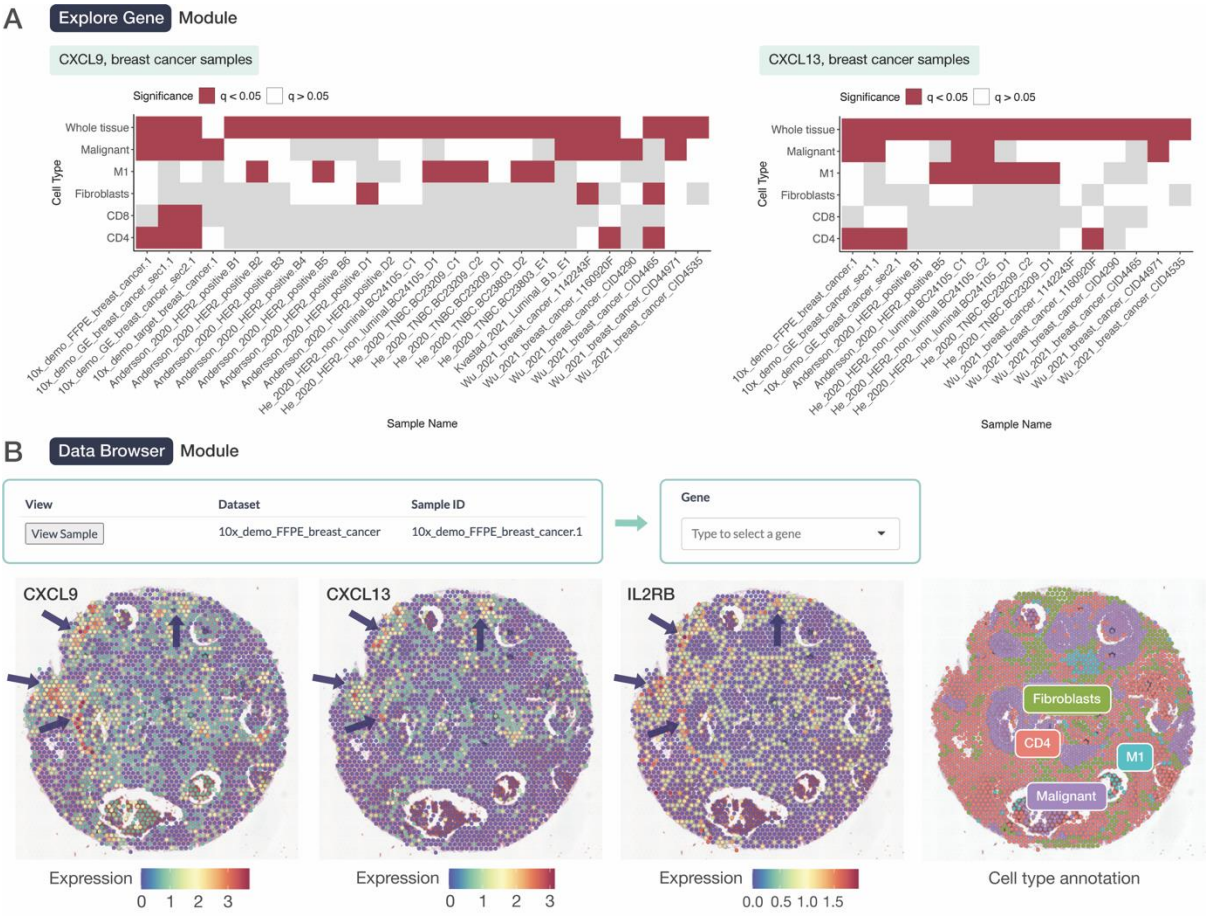

Figure 2. Examples of SOAR features exploring the spatial variability of genes in breast cancer spatial transcriptomics datasets. (A) In the "Explore Gene" module, users can visualize whether a gene is significantly spatially variable in different cell types in a heatmap. The result file can be downloaded as a tab-delimited file, and the generated figure can be downloaded in JPG and PDF formats. **(B)** In the "Data Browser" module, users can view a sample and visualize the spatial expression of a gene of interest. Black arrows indicate distinct regions of high expression of CXCL9, CXCL13, and IL2RB. q, false-discovery-rate-adjusted p-values.

Adjacency-based cell type interaction. Gene expression variation at the border between different cell types may reflect cell-cell interactions [50]. In the "Adjacency-based cell type interaction" tab, after the user selects a gene and a query cell type, the gene's differential expression between the query cell type capture locations adjacent or nonadjacent to

interacting cell types will be visualized in a heatmap. Each tile in the heatmap is coloured by the log-fold changes and sized according to their statistical significance. This allows users to explore whether the gene is involved in cell-cell interactions between the given cell type and others. Complement component 3 (C3) plays a central role in complement activation, which may aid immune surveillance against tumour cells [56,57]. In tumour microenvironments, cancer-associated fibroblasts are often the most prominent cell type [58,59] and known to express C3 [46]. As an example, we used SOAR to investigate how adjacent cell types influence C3 gene expression in cancer tissue fibroblasts. Our analysis results revealed that in cancer tissues, fibroblasts adjacent to CD4 cells have higher C3 expression levels (Figure 3A). This finding corroborates existing knowledge of the positive relationship between C3 activation and T cell infiltration in tumour tissues [60]. In contrast, cancer tissue fibroblasts adjacent to malignant cells have lower C3 levels. Previous studies have shown that cancer cells may limit C3 activation [61], which is supported by our results.

Distance-based cell type interaction. Interactions between cell types can occur beyond simple adjacency [50]. This type of long-distance communication can be characterized by regressing the gene expression levels in capture locations of a certain cell type on their Euclidean distances from other cell types. Under the “Distance-based cell type interaction” tab, users can select a gene and a query cell type, and a heatmap of the regression coefficients will be shown. The heatmap visualizes the associations between the gene’s expression in capture locations of the query cell type and their distances from potential interacting cell types, as well as whether the associations are significant. This allows users to evaluate any possible far-reaching cell-cell interactions that might not have been captured in the adjacency-based analysis. As an illustration, we studied whether the expression levels of RPLP1 in malignant capture locations were influenced by their distances from other cell types. In our analysis results, RPLP1 often had lower expression levels in malignant capture

locations when they were closer to M1 macrophages (Figure 3B). RPLP1 is a ribosomal protein gene found to be upregulated in tumour tissues [62,63] and promote cancer cell invasion [64]. On the other hand, M1 macrophages secrete proinflammatory cytokines and contribute to inflammation response against migrating tumour cells [65]. The inhibitory effect of M1 macrophages on the expression of RPLP1 in tumour cells, and in turn, on cell invasion [64], is supported by our results.

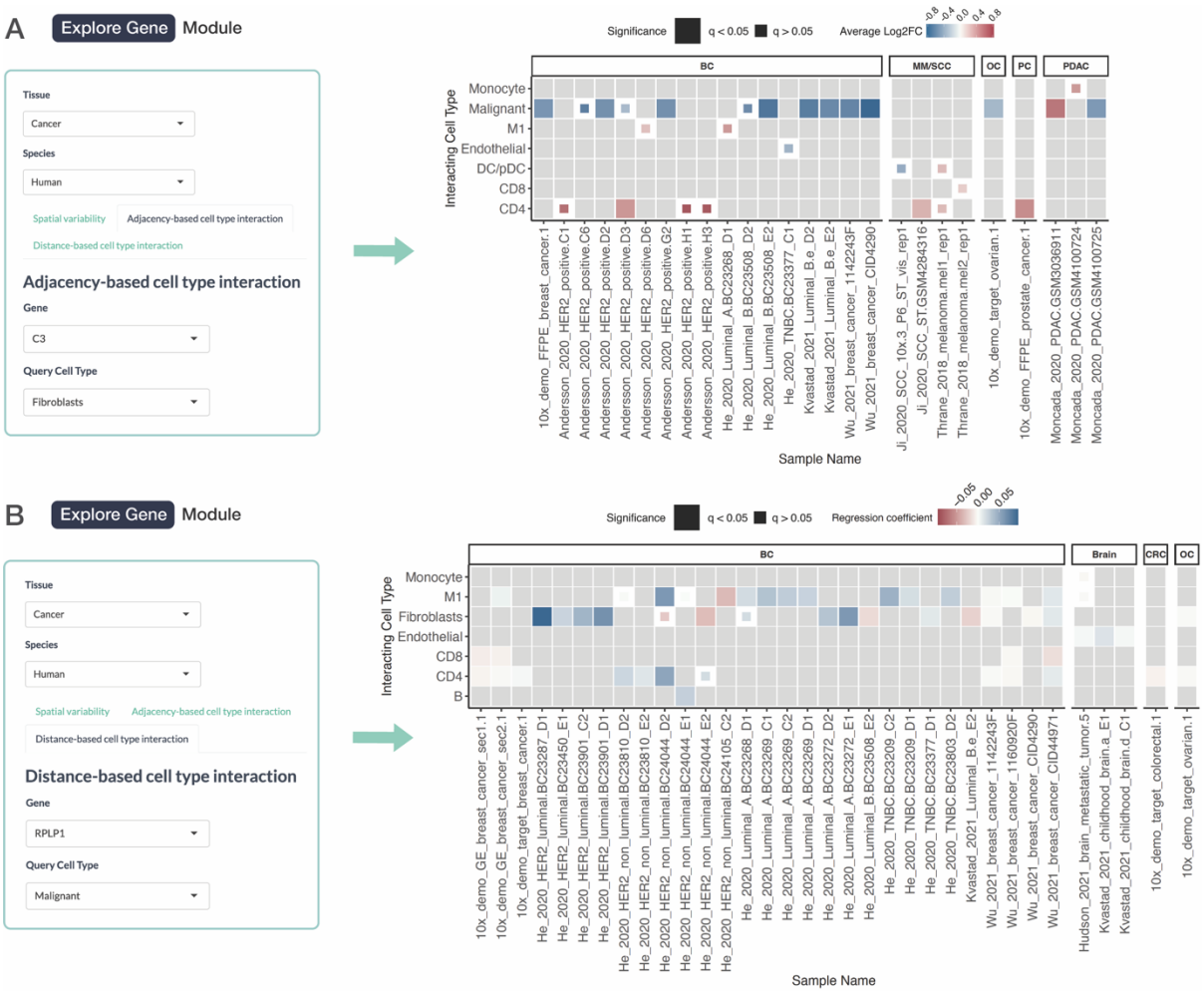

Figure 3. Examples of SOAR features exploring cell type interactions in cancer spatial transcriptomics datasets. In the "Explore Gene" module, users can interactively input a tissue type, species, gene, and query cell type and explore **(A)** adjacency-based and **(B)** distance-based cell type interactions. The results can be downloaded as a tab-delimited file, and the generated figure can be downloaded in JPG and PDF formats. **(A)** Adjacency is

defined according to the spatial coordinates of tissue capture locations, and Wilcoxon rank sum tests were used to evaluate the gene's differential expression between the query cell type capture locations adjacent or nonadjacent to interacting cell types. The log-fold changes and their significance are displayed in the heatmap. **(B)** The expression of a gene in capture locations of a query cell type is regressed on their median Euclidean distance to interacting cell types. Positive (blue) regression coefficients indicate inhibitory gene regulatory effects, and negative (red) regression coefficients suggest promotional effects. The significance levels of regression coefficients are shown in the heatmap. BC, breast cancer; CRC, colorectal cancer; Log2FC, log-fold changes; MM, melanoma; OC, ovarian cancer; PC, prostate cancer; PDAC, pancreatic ductal adenocarcinoma; q, false-discovery-rate-adjusted p-values; SCC, squamous cell carcinoma.

DISCUSSION

Spatial transcriptomics enables researchers to study gene expressions in a spatially contextualized way and hence offers rich and powerful data for a wide array of research pursuits, including biological mechanism elucidation and clinical biomarker discovery. However, due to the novel nature of this technology, accessing and utilizing published spatial transcriptomics data via existing data repositories can be quite challenging, impeding the full potential of this new and exciting technology. To facilitate future spatial transcriptomics research, we developed SOAR, a Spatial transcriptOmics Analysis Resource hosting a large number of downloadable spatial transcriptomics datasets in standardized data format. SOAR also provides a user-friendly analytic and visualization framework for visualizing spatial gene expressions, evaluating the spatial variability of genes in different tissues, and assessing possible cell type interactions. Our case studies show that users may derive biologically meaningful insights from these analysis and visualization tools. Over the coming years, we anticipate the spatial transcriptomics research community to only grow larger. SOAR will

continue to grow alongside this trend and be maintained to continuously offer useful analysis and meta-analysis tools for future researchers.

AVAILABILITY

SOAR is available at https://soar.fsm.northwestern.edu free and open to all users with no login requirement. All spatial transcriptomics datasets and analysis results can be downloaded from the data download page.

FUNDING

This work was supported by the National Institutes of Health R01LM013337, 5UL1TR001422.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1 Zeng, Z., Li, Y., Li, Y. & Luo, Y. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biol.* **23**, 1-23 (2022).
2 Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).
3 Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature biotechnology* **38**, 333-342 (2020).
4 He, B. *et al.* Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature biomedical engineering* **4**, 827-834 (2020).
5 Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nature genetics* **53**, 1334-1347 (2021).
6 Ji, A. L. *et al.* Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 497-514. e422 (2020).
7 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 research* **78**, 5970-5979 (2018).
8 Berglund, E. *et al.* Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature communications* **9**, 1-13 (2018).
9 Andersson, A. *et al.* Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nature communications* **12**, 1-14 (2021).
10 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).
11 Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342-357 (2016).
12 Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience* **24**, 425-436 (2021).
13 Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature methods* **15**, 932-935 (2018).
14 Zhu, Q., Shah, S., Dries, R., Cai, L. & Yuan, G.-C. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nature biotechnology* **36**, 1183-1190 (2018).
15 Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241. e3226 (2021).
16 Ratz, M. *et al.* Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. *Nature neuroscience*, 1-10 (2022).
17 Ortiz, C. *et al.* Molecular atlas of the adult mouse brain. *Science advances* **6**, eabb3446 (2020).
18 Navarro, J. F. *et al.* Spatial transcriptomics reveals genes associated with dysregulated mitochondrial functions and stress signaling in Alzheimer disease. *Iscience* **23**, 101556 (2020).
19 Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
20 Joglekar, A. *et al.* A spatially resolved brain region-and cell type-specific isoform atlas of the postnatal mouse brain. *Nature Communications* **12**, 1-16 (2021).
21 Hasel, P., Rose, I. V., Sadick, J. S., Kim, R. D. & Liddelow, S. A. Neuroinflammatory astrocyte subtypes in the mouse brain. *Nature neuroscience* **24**, 1475-1487 (2021).
22 Chen, R. *et al.* Decoding molecular and cellular heterogeneity of mouse nucleus accumbens. *Nature Neuroscience* **24**, 1757-1771 (2021).
23 Chen, W.-T. *et al.* Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* **182**, 976-991. e919 (2020).
24 Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature biotechnology* **39**, 313-319 (2021).
25 Fawkner-Corbett, D. *et al.* Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810-826. e823 (2021).

26 Liu, Y. *et al.* High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665-1681. e1618 (2020).
27 Lohoff, T. *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nature biotechnology* **40**, 74-85 (2022).
28 Sanchez-Ferras, O. *et al.* A coordinated progression of progenitor cell states initiates urinary tract development. *Nature communications* **12**, 1-16 (2021).
29 D'Adamo, G. L., Widdop, J. T. & Giles, E. M. The future is now? Clinical and translational aspects of "Omics" technologies. *Immunol. Cell Biol.* **99**, 168-176 (2021).
30 Luo, Y. *et al.* A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia. *Nat. Med.* **26**, 1375–1379, doi:10.1038/s41591-020-1007-0 (2020).
31 Kline, A. *et al.* Multimodal Machine Learning in Precision Health. *arXiv preprint arXiv:2204.04777* (2022).
32 Luo, Y. *et al.* Integrating Hypertension Phenotype and Genotype with Hybrid Non-negative Matrix Factorization. *Bioinformatics* **35**, 1395-1403, doi:10.1093/bioinformatics/bty804 (2019).
33 Canuel, V., Rance, B., Avillach, P., Degoulet, P. & Burgun, A. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. *Briefings in bioinformatics* **16**, 280-290 (2015).
34 Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
35 Moffitt, J. R. *et al.* High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proceedings of the National Academy of Sciences* **113**, 14456-14461 (2016).
36 Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nature methods* **11**, 360-361 (2014).
37 Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235-239 (2019).
38 Fan, Z., Chen, R. & Chen, X. SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Research* **48**, D233-D237 (2020).
39 Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nature Methods*, 1-13 (2022).
40 Maniatis, S. *et al.* Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89-93 (2019).
41 Qian, J. *et al.* A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell research* **30**, 745-762 (2020).
42 Ren, X. *et al.* COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* **184**, 1895-1913. e1819 (2021).
43 Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902. e1821 (2019).
44 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome biology* **20**, 1-15 (2019).
45 Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature immunology* **20**, 163-172 (2019).
46 Sun, D. *et al.* TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic acids research* **49**, D1420-D1430 (2021).
47 Miller, J. A. *et al.* Common cell type nomenclature for the mammalian brain. *Elife* **9**, e59928 (2020).
48 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nature methods* **14**, 1083-1086 (2017).
49 Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biology* **22**, 1-25 (2021).
50 Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics* **22**, 71-88 (2021).
51 Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nature methods* **15**, 343-346 (2018).

52 Bronger, H. *et al.* CXCL9 and CXCL10 predict survival and are regulated by cyclooxygenase inhibition in advanced serous ovarian cancer. *British journal of cancer* **115**, 553-563 (2016).
53 Fukuda, Y. *et al.* Endogenous CXCL9 affects prognosis by regulating tumor‐infiltrating natural killer cells in intrahepatic cholangiocarcinoma. *Cancer science* **111**, 323-333 (2020).
54 Pelka, K. *et al.* Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734-4752. e4720 (2021).
55 Setrerrahmane, S. & Xu, H. Tumor-related interleukins: old validated targets for new anti-cancer drug development. *Molecular cancer* **16**, 1-17 (2017).
56 Afshar-Kharghan, V. The role of the complement system in cancer. *The Journal of clinical investigation* **127**, 780-789 (2017).
57 Shu, C. *et al.* C3a-C3aR signaling promotes breast cancer lung metastasis via modulating carcinoma associated fibroblasts. *Journal of Experimental & Clinical Cancer Research* **39**, 1-14 (2020).
58 Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nature medicine* **19**, 1423-1437 (2013).
59 Sahai, E. *et al.* A framework for advancing our understanding of cancer-associated fibroblasts. *Nature Reviews Cancer* **20**, 174-186 (2020).
60 Lin, K. *et al.* Complement component 3 is a prognostic factor of non-small cell lung cancer. *Molecular medicine reports* **10**, 811-817 (2014).
61 Kleczko, E. K., Kwak, J. W., Schenk, E. L. & Nemenoff, R. A. Targeting the complement pathway as a therapeutic strategy in lung cancer. *Frontiers in immunology* **10**, 954 (2019).
62 Artero-Castro, A. *et al.* Rplp1 bypasses replicative senescence and contributes to transformation. *Experimental cell research* **315**, 1372-1383 (2009).
63 Artero-Castro, A. *et al.* Expression of the ribosomal proteins Rplp0, Rplp1, and Rplp2 in gynecologic tumors. *Human pathology* **42**, 194-203 (2011).
64 He, Z. *et al.* RPLP1 promotes tumor metastasis and is associated with a poor prognosis in triple-negative breast cancer patients. *Cancer cell international* **18**, 1-10 (2018).
65 Lin, Y., Xu, J. & Lan, H. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. *Journal of hematology & oncology* **12**, 1-16 (2019).