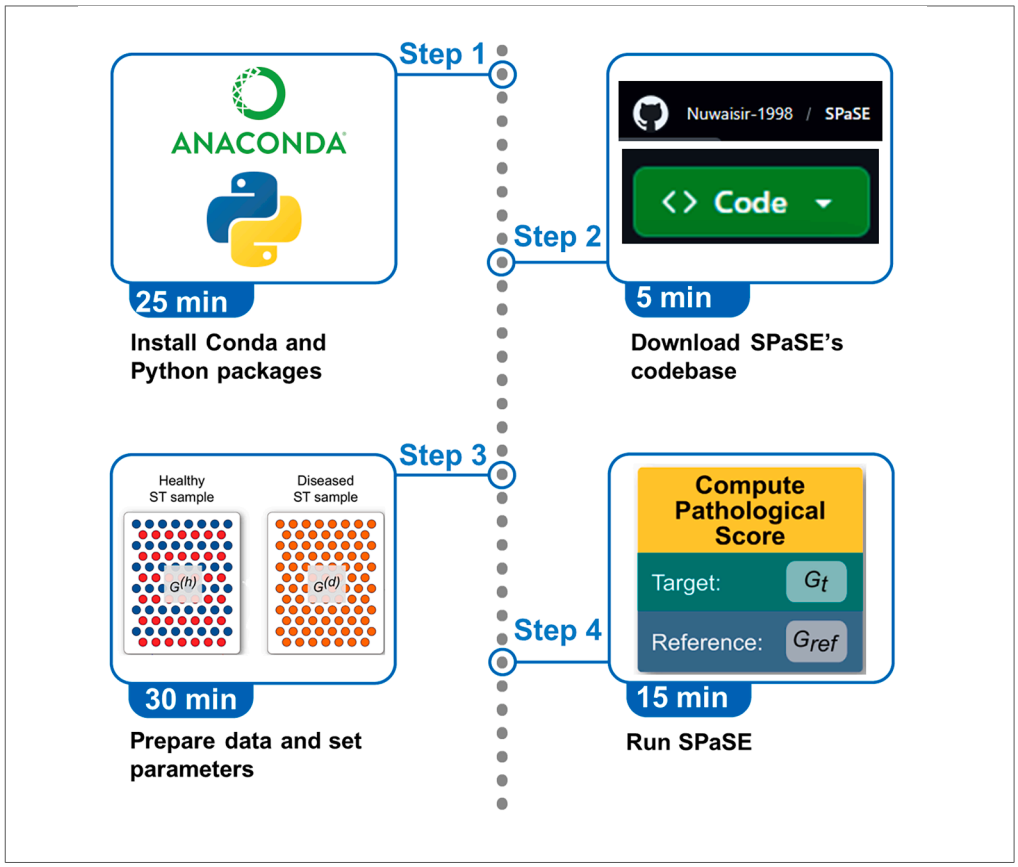
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for spatially resolved pathology scores using optimal transport on spatial transcriptomics data



Spatial transcriptomics (ST) integrates gene expression with spatial context to study tissue pathology. We present SPaSE (spatially resolved pathology score), a protocol to quantify pathology by computing an optimal transport plan between ST spots of the diseased and the corresponding healthy tissue. We describe steps for assigning pathology scores to each spot and developing support vector regression models to link gene expression with pathological impact. We demonstrate the utility of SPaSE across multiple human and mouse ST datasets.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Mohammad Nuwaisir Rahman, James F. Martin, M Saifur Rahman, Md. Abul Hassan Samee

nuwaisir1998@gmail.com (M.N.R.)
mrahman@cse.buet.ac.bd (M.S.R.)
samee@bcm.edu (M.A.H.S.)

Highlights

Step-by-step instructions to install Conda and set up the SPaSE codebase

Guidelines for data preparation and parameter configuration for SPaSE

Instructions to execute SPaSE and locate generated results

Rahman et al., STAR Protocols 7, 104584
June 19, 2026 Published by Elsevier Inc.
https://doi.org/10.1016/j.xpro.2026.104584

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for spatially resolved pathology scores using optimal transport on spatial transcriptomics data

Mohammad Nuwaisir Rahman,[1,2,3,6,*] James F. Martin,[4,5] M Saifur Rahman,[1,*] and Md. Abul Hassan Samee[4,7,*]

[1]Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

[2]Computer Science and Engineering, BRAC University, Dhaka 1212, Bangladesh

[3]Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119077, Singapore

[4]Integrative Physiology, Baylor College of Medicine, Houston, TX 77030, USA

[5]Texas Heart Institute, Houston, TX 77030, USA

[6]Technical contact

[7]Lead contact

*Correspondence: nuwaisir1998@gmail.com (M.N.R.), mrahman@cse.buet.ac.bd (M.S.R.), samee@bcm.edu (M.A.H.S.)
https://doi.org/10.1016/j.xpro.2026.104584

SUMMARY

Spatial transcriptomics (ST) integrates gene expression with spatial context to study tissue pathology. We present SPaSE (spatially resolved pathology score), a protocol to quantify pathology by computing an optimal transport plan between ST spots of the diseased and the corresponding healthy tissue. We describe steps for assigning pathology scores to each spot and developing support vector regression models to link gene expression with pathological impact. We demonstrate the utility of SPaSE across multiple human and mouse ST datasets.

For complete details on the use and execution of this protocol, please refer to Rahman et al.[1]

BEFORE YOU BEGIN

The protocol below describes the specific steps of using SPaSE to generate pathology scores of an ST tissue sample. To calculate the pathology scores of the spots of an ST sample, SPaSE generates an optimal transport plan between the spots of a healthy and the target pathological sample. Based on the learned optimal transport plan, SPaSE assigns a score to each of the ST spots of the target pathological sample. We tested SPaSE on mouse heart post-myocardial infarction,[2] mouse traumatic brain injury,[3] and mouse Duchenne muscular dystrophy[4] ST samples. In all of the cases, spots in the pathological regions show higher pathology scores, confirming SPaSE's broad applicability.

Innovation

This protocol details the implementation of SPaSE, an optimal transport-based computational method for quantifying spatially variable pathological effects in ST data. Existing spatial domain detection tools identify transcriptionally homogeneous regions but fail to quantify the severity or gradients of pathological impact across tissues. SPaSE addresses this critical gap by introducing a principled framework that computes real-valued pathology scores for each spatial location, enabling quantitative, spatially continuous assessment of disease burden. The core innovation of SPaSE lies in its optimal transport formulation, which models the transcriptomic difference between control and diseased samples as a transport cost. This allows for measurement of local

transcriptomic differences while preserving spatial topology. SPaSE further incorporates a null distribution derived from control samples to compute statistical significance (p values) of pathology scores, offering a rigorous probabilistic basis for spatial pathology inference. Unlike prior correlation- or clustering-based methods, SPaSE yields interpretable, continuous pathology maps that can be directly linked to downstream transcriptional mechanisms through regression or enrichment analyses. Its modular design and generalizable statistical framework make it applicable across diverse tissue types, disease models, and species. Together, these innovations position SPaSE as a powerful and extensible computational approach for quantifying and interpreting spatial pathology from ST data.

Install Miniconda and Python packages

Timing: 25 min

This section describes the process of downloading Miniconda for Linux, Windows, or Mac.

1. Download and install Miniconda.
   a. In case of Linux (×86 architecture),
      i. Download Minicoda using the following command in terminal:
      `>wget` https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
      ii. Install Miniconda by running:
      `>bash ~/Miniconda3-latest-Linux-x86_64.sh`
   b. In case of Windows,
      i. Download Miniconda using the following command in command prompt:
      `>curl` https://repo.anaconda.com/miniconda/Miniconda3-latest-Windows-x86_64.exe `--output .\Downloads\Miniconda3-latest-Windows-x86_64.exe`
      ii. Navigate to the Downloads folder and run the installer.
   c. In case of Mac,
      i. Download Minicoda using the following command in terminal:
      `>curl -O` https://repo.anaconda.com/miniconda/Miniconda3-latest-MacOSX-arm64.sh
      ii. Install Miniconda by running:
      `>bash ~/Miniconda3-latest-MacOSX-arm64.sh`

Download SPaSE's codebase

Timing: 5 min

This section describes the process of downloading SPaSE's codebase.

2. Download SPaSE's codebase.
   a. Open a browser and search for https://github.com/Nuwaisir-1998/SPaSE.
   b. Under the 'Code' dropdown, select 'Download Zip'.
   c. Extract the downloaded zip file in your desired location. A directory named 'SPaSE' will be extracted.
   d. Open a terminal/command prompt and navigate to 'SPaSE' directory.
   e. Run the following command to create the conda environment for SPaSE:
   `>conda env create -f environment.yml`
   Check the troubleshooting section if the environment is not created properly.
   f. Activate the environment using the following command:
   `>conda activate spase`

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Spatial transcriptomics (Visium, 10× Genomics) data of healthy and injured mouse heart. Injuries include MI (permanent ligation of the LAD coronary artery), I/R-30 (30 minutes of ischemia followed by reperfusion), transaortic constriction, isoproterenol injection, and traumatic needle injection.	Calcagno et al.	GEO: GSE214611
Spatial transcriptomics (Visium, 10× Genomics) data of Duchenne mouse models	Heezen et al.	GEO: GSE199659
Spatial transcriptomics data of mouse traumatic brain injury	Pham et al.	GEO: GSE236171
Spatial multi-omic map of human myocardial infarction	Kuppe et al.	DOI: https://doi.org/10.5281/zenodo.6578046
Software and algorithms		
Miniconda	Anaconda	http://anaconda.com/docs/getting-started/miniconda/main
Python v3.8.13	Python Software Foundation	https://www.python.org/downloads/release/python-3813/
POT v0.8.2	Flamary et al.	https://PythonOT.github.io/
SHAP v0.42.0	Su-In Lee's lab at the University of Washington, and Microsoft Research.	https://shap.readthedocs.io/en/latest/
Other		
Original codes	This manuscript	DOI: https://doi.org/10.5281/zenodo.14941956

STEP-BY-STEP METHOD DETAILS

Prepare data and set parameters

Timing: 30 min

This section outlines the data preparation steps required to run SPaSE, along with the configuration of necessary parameters and hyperparameters, and a detailed description of each hyperparameter.

1. Obtain spatial transcriptomics data in AnnData format.
   a. Ensure you have both healthy (control) and diseased ST samples.
   b. Data should be in.h5ad format with gene expression counts and spatial coordinates.
   c. Verify that spatial coordinates are stored in the.obsm['spatial'] attribute.
2. Select hyperparameters (alpha, and lambda) using grid search.
   a. Navigate to Workspace/SPaSE/src/
   b. Run the following command:

```
>python perturb.py --adata_path [path to healthy sample's .h5ad file]
--adata_save_path [path to save the perturbed sample]
```

Note: The file name of each perturbed sample should follow the format of the corresponding healthy sample's name, appended with "_[x]" to indicate the x-th perturbation.

   i. Example command:

```
>python perturb.py --adata_path ./Data/adata.h5ad --adata_save_path
./Data/adata_1.h5ad
```

   c. Generate 3 different perturbed healthy samples using this procedure.
   d. Update the set of hyperparameters to be used by modifying *alphas* and *lambda_sinkhorns* list inside *run_grid_search.py*. *alpha* controls the relative importance of the spatial component in the SPaSE objective.

Note: Increasing alpha places greater emphasis on spatial consistency, making the resulting scores more spatially aware by favoring solutions that better respect spatial alignment. Conversely, smaller values of α reduce the influence of spatial information and shift the model toward relying more heavily on expression-based similarity. The parameter lambda scales the regularization term in the Sinkhorn optimization. Larger values of lambda impose stronger regularization, leading to smoother and more stable transport plans, while smaller values allow for sharper, more localized matches at the cost of increased sensitivity to noise. Running time remains largely stable across different alpha and lambda values, though lower lambda may slightly increase computation time due to slower convergence requiring additional iterations.

e. Run grid search using the following command:
```
>python run_grid_search.py --dataset [Dataset name, provide a name of your choice] --adata_left_path [.h5ad file path of one healthy sample] --adata_healthy_right_path [.h5ad file path of another healthy sample] --adata_right_path [.h5ad file path of the perturbed sample].
```
   i. Example command:
```
>python run_grid_search.py --dataset Mouse_heart --adata_left_path ./Data/adata.h5ad --adata_healthy_right_path ./Data/adata_healthy_right.h5ad --adata_right_path ./Data/adata_1.h5ad
```
f. Find the best hyperparameters using the following command:
```
>python hyperparameter_selection.py --dataset [Dataset name] --sample [Sample filename without the extension].
```
   i. Example command:
```
>python hyperparameter_selection.py --dataset Mouse_heart --sample adata
```

Run SPaSE

Timing: 15 min

This section explains how to run SPaSE, describes the parameters required for execution, and specifies where the results are stored.

3. Run SPaSE with your parameters.
   a. Navigate to Workspace/SPaSE/src/.
   b. Run the following command:
```
>python run.py --dataset_name [user input] --healthy [user input] --healthy_right [user input] --diseased [user input] --alpha [user input] --lambda_sinkhorn [user input]
```
      i. Description of the parameters:

dataset_name [type: string].

healthy [type: string; the path of the.h5ad file that would be used as reference].

diseased [type: string; the path of the.h5ad file that would be used as target].

healthy_right [Optional parameter; type: string; the path of the.h5ad file representing a second healthy sample. If it is not provided, the healthy sample will be decomposed into two and those two samples will be treated as two healthy samples.].

alpha [type: double; hyperparameter].

lambda_sinkhorn [type: double; hyperparameter].

Check the troubleshooting section if this results in any error.

> ***Note:*** The results will be stored inside Workspace/SPaSE/results/ under the name you provide in the dataset variable.

EXPECTED OUTCOMES

SPaSE generates spatially resolved pathology scores for each spot in your diseased tissue sample. These scores typically range from 0 to 100, with higher values indicating greater pathological impact. In our validation studies using mouse heart post-myocardial infarction data, we observed clear delineation of infarct zones, border zones, and remote zones, which matched delineations by Calcagno et al.

LIMITATIONS

A key limitation of SPaSE lies in its reliance on the assumption that tissue sections are fully alignable. At present, SPaSE is not well-suited for scenarios involving partially overlapping slides, which poses a significant challenge for applying the method to complex human tissue data where perfect alignment is rarely achievable. This limitation directly affects its applicability to real-world datasets, although the current support vector regression (SVR) model provides a workaround by predicting pathology scores from gene expressions.

TROUBLESHOOTING

Problem 1 (related to step: Install Conda and Python packages)

Conda environment may not be created properly because of broken Conda installation file.

Potential solution

Verify Conda is installed and internet connection is stable. Try creating environment manually with individual packages with either Conda, Mamba, or pip package manager. The list of required packages can be found inside *requirements.txt* file.

Problem 2 (related to step: Prepare data and set parameters)

Missing spatial coordinates of the ST tissue samples.

Potential solution

Verify if the spatial coordinates are stored under the *.obsm* using key *'spatial'* in the anndata object. Let, *adata* be the anndata object, then *adata.obsm['spatial']* should contain the spatial information.

Problem 3 (related to step: Prepare data and set parameters)

A data format that differs from the one currently used by SPaSE.

Potential solution

SpatialData[5] package can be used to transform among various types of spatial transcriptomics data formats.

Problem 4 (related to step: Run SPaSE)

Memory errors during computation.

Potential solution

Memory overflow can occur when VRAM is insufficient during GPU computations. To prevent this, consider using the CPU to store data in system RAM instead of relying solely on VRAM. To use CPU, set *use_gpu* to *0* in run.py file.

Problem 5 (related to step: Run SPaSE)

Optimal transport algorithm fails to converge.

Potential solution

Adjust *alpha* and *lambda_sinkhorn* parameters. Try *alpha* from the range [0.01, 0.1], *lambda_sinkhorn* from the range [0.001, 0.1].

Problem 6 (related to step: Run SPaSE)

Extremely long running time (more than 1 hour to find the pathology scores of a sample).

Potential solution

Executing instructions using CPU may cause long running time. It is preferable to use GPU to lessen the running time significantly. To use GPU, set *use_gpu* to 1 *in run.py file*.

Problem 7 (related to step: Run SPaSE)

If there is only one healthy sample available, and it is not Visium, the synthesis process cannot be applied.

Potential solution

Randomly split the spots/cells into two halves and use them to create two synthetic healthy samples.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Md. Abul Hassan Samee (samee@bcm.edu).

Technical contact

Technical questions on executing this protocol should be directed to and will be answered by the technical contact, Mohammad Nuwaisir Rahman (nuwaisir1998@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- ST data of mouse hearts can be found in the Gene Expression Omnibus (GEO) under accession no. GSE214611. The raw data files for the mouse DMD samples can be accessed in the GEO under accession no. GSE199659. The processed data files are available on Zenodo at https://doi.org/10.5281/zenodo.7401196. The raw and processed sequencing data of the mouse TBI sample have been deposited in the GEO under accession no. GSE236171 and are publicly available. The spatial transcriptomics of human heart samples can be accessed on CellxGene at https://cellxgene.cziscience.com/collections/8191c283-0816-424b-9b61-c3e1d6258a77 and in the Zenodo data archive at https://zenodo.org/record/6578047.
- All original code has been deposited at https://zenodo.org/records/14941956 and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

M.N.R. was supported by the ICT Fellowship at the master's level in the fiscal year 2022–2023 for research in ICT sector, administered by the ICT Division, Government of People's Republic of Bangladesh. M.S.R. was partially supported by a Basic Research Grant from BUET. J.F.M. was supported by the National Institutes of Health (HL169511, HL171574, and HL118761 to J.F.M.) and the Vivian L. Smith Foundation. M.A.H.S. was supported by the National Institutes of Health (AG081192).

AUTHOR CONTRIBUTIONS

Conceptualization, M.A.H.S.; algorithm design (the SPaSE algorithm and downstream analyses), M.N.R., M.S.R., and M.A.H.S.; implementation (the SPaSE algorithm and downstream analyses), M.N.R.; result interpretation, M.N.R., J.F.M., M.S.R., and M.A.H.S.; manuscript writing, M.N.R. with inputs from M.S.R., J.F.M., and M.A.H.S.; all authors edited the manuscript.

DECLARATION OF INTERESTS

J.F.M. is a co-founder and owns shares in YAP Therapeutics.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, we used ChatGPT to improve the readability and language of the manuscript. After using this tool, we reviewed and edited the content as necessary and take full responsibility for the final content of the published article.

REFERENCES

1. Rahman, M.N., Abrar, M.A., Shaw, V.R., Martin, J.F., Saifur Rahman, M., and Samee, M.A.H. (2025). SPaSE: Spatially resolved pathology scores using optimal transport on spatial transcriptomics data. Cell Syst. *16*, 101301.

2. Calcagno, D.M., Taghdiri, N., Ninh, V.K., Mesfin, J.M., Toomu, A., Sehgal, R., Lee, J., Liang, Y., Duran, J.M., Adler, E., et al. (2022). Single-cell and spatial transcriptomics of the infarcted heart define the dynamic onset of the border zone in response to mechanical destabilization. Nat. Cardiovasc. Res. *1*, 1039–1055.

3. Pham, D., Tan, X., Xu, J., Grice, L.F., Lam, P.Y., Raghubar, A., Vukovic, J., Ruitenberg, M.J., and Nguyen, Q. (2020). stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. bioRxiv. https://doi.org/10.1101/2020.05.31.125658. https://www.biorxiv.org/content/early/2020/05/31/2020.05.31.125658.full.pdf.

4. Heezen, L.G.M., Abdelaal, T., van Putten, M., Aartsma-Rus, A., Mahfouz, A., and Spitali, P. (2023). Spatial transcriptomics reveal markers of histopathological changes in duchenne muscular dystrophy mouse models. Nat. Commun. *14*, 4909.

5. Marconato, L., Palla, G., Yamauchi, K.A., Virshup, I., Heidari, E., Treis, T., Vierdag, W.M., Toth, M., Stockhaus, S., Shrestha, R.B., et al. (2025). SpatialData: an open and universal data framework for spatial omics. Nat. Methods *22*, 58–62.