

SPACESEQUENT: A UNIFIED PIPELINE FOR SPATIAL TRANSCRIPTOMICS DATA ANALYSIS

Yu H. Sun^{1,*,#}, Sarbottam Piya^{1,#}, Zhengyu Ouyang^{2,#}, Yirui Chen¹, Jake Gagnon³, Shaolong Cao¹, Haotian Zhang¹, Baobao Song¹, Jing Zhu¹, Khyati Chandratre¹, Houlin Yu¹, Wenxing Hu¹, Matthew Ryals⁴, Fergal Casey¹, Dann Huh¹, Baohong Zhang¹

¹ Research, Biogen Inc., Cambridge, MA, 02142, USA

² Data Science, BioInfoRx Inc., Madison, WI, 53719, USA

³ Biostatistics, Biogen Inc., Cambridge, MA, 02142, USA

⁴ Cencora Global Consulting Services US Corporation, Conshohocken, PA, 19428, USA

Contribute equally

Yu H. Sun: <https://orcid.org/0000-0003-0333-2898>

Sarbottam Piya: <https://orcid.org/0000-0002-1070-8008>

Yirui Chen: <https://orcid.org/0000-0003-1522-8908>

Shaolong Cao: <https://orcid.org/0000-0002-4443-9424>

Haotian Zhang: <https://orcid.org/0009-0002-2309-7936>

Baobao Song: <https://orcid.org/0000-0002-0062-2710>

Khyati Chandratre: <https://orcid.org/0000-0003-3698-3670>

Houlin Yu: <https://orcid.org/0000-0002-5374-2783>

Wenxing Hu: <https://orcid.org/0000-0003-2477-6383>

Fergal Casey: <https://orcid.org/0000-0002-6781-8670>

Baohong Zhang: <https://orcid.org/0000-0003-4735-4212>

Corresponding Author:

yuhenry.sun@biogen.com

To Cite This Article: SPACESEQUENT: A UNIFIED PIPELINE FOR SPATIAL TRANSCRIPTOMICS DATA ANALYSIS (Y. H. Sun, S. Piya, Z. Ouyang, Y. Chen, J. Gagnon, S. Cao, H. Zhang, B. Song, J. Zhu, K. Chandratre, H. Yu, W. Hu, M. Ryals, F. Casey, D. Huh, & B. Zhang, Trans.). (2026). Journal of Advance Research in Pharmacy and Biological Science (ISSN 2208-2360), 12(1). <https://doi.org/10.61841/nn-pbs-12-1-1>

ABSTRACT

Spatial transcriptomics has emerged as one of the most powerful tools for gaining biological insights, enabling researchers to uncover intricate relationships between gene expression patterns and tissue architecture. Recent advances in the field have resulted in a variety of new platforms, including Visium, Visium HD, and Xenium from 10x Genomics, as well as GeoMx and CosMx from NanoString Technologies, which has now been acquired by the Bruker Corporation. However, the existence of diverse spatial transcriptomics platforms and various data formats poses challenges in standardizing data analysis. Thus, there remains a critical gap in the availability of a comprehensive pipeline capable of conducting end-to-end analysis that is necessary to extract biological insights from multiple spatial transcriptomics platforms. Here, we present SpaceSequest, a tailored pipeline that utilizes cutting-edge computational methodologies to conduct a thorough analysis, enabling the extraction of crucial biological insights from five major spatial transcriptomics technologies. SpaceSequest performs (1) standardized quality control and general data processing, (2) key analyses customized for each spatial platform, (3) automated cell type annotation and deconvolution, and (4) high-quality figure and analysis result generation. In addition, SpaceSequest allows for smooth integration with cellxgene VIP and Quickomics for user-friendly data access and interactive visualization. In short, SpaceSequest is a unified and comprehensive pipeline designed for the analysis, visualization, and publication of spatial transcriptomics data from various platforms. The source code is available at <https://github.com/interactivereport/SpaceSequest>. To facilitate seamless installation and usage, we have also created a detailed Bookdown tutorial that can be accessed through <https://interactivereport.github.io/SpaceSequest/tutorial/docs/index.html>.

KEYWORDS: Spatial transcriptomics, Visium, Visium HD, Xenium, GeoMx, CosMx, single-cell RNA-seq, deconvolution

© Author(s) 2026. This work is distributed under the Creative Commons BY- 4.0 license:

<https://creativecommons.org/licenses/by/4.0/>

1. INTRODUCTION

The advent of single-cell RNA sequencing (scRNA-seq) has revolutionized our understanding of cellular heterogeneity and gene expression dynamics. It represents a significant advancement from bulk RNA-seq, which only provides average expression profiles from mixed cell populations, thereby obscuring the underlying cellular diversity¹⁻³. Although scRNA-seq generates invaluable insights into cellular heterogeneity, it lacks spatial information, as it dissociates cells from their original tissue context^{4,5}. Understanding the intricate spatial organization of gene expression within complex tissues and organs is crucial for unraveling the mysteries of biology⁶⁻⁸. Spatial transcriptomics (ST) has emerged as a transformative technology that quantifies RNA within intact tissue sections, enabling researchers to explore gene expression patterns within their native spatial context^{6,9-12}.

ST can provide precise insights into disease progression by discerning heterogeneity within diseased tissue, which is powerful for studying complex diseases such as tumor and neurological disorders¹³⁻¹⁷. For example, using ST data, Berglund et al. not only distinguished healthy and diseased areas but were able to identify re-stratified tumor heterogeneity by analyzing gene expression gradients within the tumor microenvironment, specifically in the stroma adjacent to the tumor area¹⁸. These advancements have unlocked new avenues for addressing critical biological inquiries. First, ST plays a pivotal role in characterizing the cell-type composition of tissues. While definitions are often derived from large-scale single-cell RNA-seq or epigenetics datasets and projected onto ST data, the technology also empowers the direct generation of cell-type definitions from the spatial transcriptomics data itself^{19,20}. This application has been widely utilized, resulting in the creation of compositional atlases for a diverse range of tissues, including the nervous system²¹⁻²⁴, human kidney²⁵⁻²⁷, heart^{28,29}, testes³⁰, and lung^{31,32}. Second, ST facilitates the exploration of cellular interactions, revealing the rules and patterns governing spatial covariation among distinct cell types. For instance, studies examining the mouse visual cortex identified a preferential spatial proximity of inhibitory neuron subtypes, shedding light on their organizational principles³³. Similar proximity analyses have been instrumental in elucidating gene expression patterns linked to Alzheimer's disease and in characterizing histopathological responses to traumatic brain injury³⁴⁻³⁷. Finally, ST's transcriptome-wide data on in situ gene expression empowers the elucidation of molecular interactions between tissue components⁷. Through the identification of ligand-receptor pairs among cell types at varying spatial proximities, we can discern the intricate communication networks that underlie phenomena such as tumor interactions with the microenvironment, immune infiltrates in tissues, and the establishment of developmental gradients^{38,39}.

There are two distinct ST methods: (i) sequencing-based ST, and (ii) imaging-based ST. In the sequencing-based ST method, a DNA-barcoded surface is used to capture RNA molecules from the intact tissue section. These captured RNA molecules are then subjected to next-generation sequencing to determine their respective gene identities⁶. By contrast, in imaging-based methods, multiple mRNA transcripts are simultaneously detected within intact tissue using fluorescence in situ hybridization (FISH) or direct in-situ sequencing^{6,12}. The imaging-based methods, such as 10x Genomics Xenium and NanoString CosMx, are suitable for hypothesis testing as these methods allow transcriptomic profiling at single cell resolution albeit some platforms can only profile a limited number of target genes. Conversely, sequencing-based methods which produce an unbiased, whole-transcriptome profile, are suitable for hypothesis generation. Typical methods include 10x Genomics Visium, Visium HD, and NanoString GeoMx with sequencing readout. However, transcriptomics profiles obtained from these methods are not directly resolved at single cell resolution.

One critical drawback of current grid-based ST profiling methods (Visium, Slide-seq^{38,40}, and HDST)⁴¹ is their inability to achieve single-cell resolution. These techniques either combine multiple cells or fractions of multiple cells at each tissue location, limiting the ability to discern individual cell contributions. Nevertheless, this limitation can be addressed through computational integration, leveraging the reference transcriptome signatures of cell types obtained from coupled single-cell RNA sequencing (scRNA-seq) profiles. By integrating ST data with scRNA-seq data, researchers can deconvolute cell type composition and retrieve further information from the spatially resolved gene expression patterns^{42,43}.

The landscape of spatial transcriptomics data analysis has seen the development of several data analysis pipelines, including ST Pipeline⁴⁴ and Spacemake⁴⁵. ST Pipeline primarily focuses on the essential tools and scripts required to process raw files derived from Visium data in FASTQ format, providing datasets primed for downstream analysis⁴⁴. In contrast, Spacemake represents a more comprehensive solution, offering a wider range of functionalities compared to ST Pipeline. This tool is adept at handling spatial transcriptomic data from diverse platforms, such as Visium, Slide-seq, Scope, etc⁴⁵. Another distinctive feature of Spacemake is its utilization of NovoSpaRc⁴⁶ that allows mapping snRNASeq data to spatial coordinates⁴⁵. Recently developed pipelines are focused on specific tasks, for example, PIPEFISH⁴⁷ was designed for FISH spatial transcriptomics, ENACT was specifically designed for Visium HD data analysis⁴⁸, while Panpipes⁴⁹ was created for spatial multi-omics analysis. Although these existing pipelines are useful, their capabilities fall short in extracting comprehensive and meaningful biological insights from diverse spatial transcriptomics datasets.

Here we introduce SpaceSequest, our comprehensive spatial transcriptomics data analysis pipeline tailored for the analysis of several major platforms, spanning 10X Genomics Visium, Visium HD, Xenium, as well as NanoString GeoMx and CosMx platforms. By utilizing cutting-edge computational methodologies, SpaceSequest streamlines a series of essential

analyses to extract meaningful biological insights. Compatible with diverse input formats, the pipeline seamlessly navigates through critical steps including clustering, deconvolution, cell type annotation, and differential expression analysis. This integrated approach equips researchers with a powerful toolkit to dissect the spatially defined molecular landscape, facilitating a deeper understanding of complex biological systems and their microenvironmental interactions.

2. METHODS

2.1 GENERAL DATA PROCESSING

SpaceSquest supports various input formats and processes the data through a series of automated steps, leveraged by multiple well-designed packages including Seurat 5⁵⁰ and anndata⁵¹. For sequencing-based strategies (10x Genomics Visium, Visium HD, and NanoString GeoMx with sequencing readouts), FASTQ processing is required before running SpaceSquest. The input files for running Visium and Visium HD workflows, including fundamental quality metrics and the unique molecular identifiers (UMI) count matrix, can be generated through 10x Genomics Space Ranger software. Since our pipeline does not rely on gene annotation or species-specific details, users are required to accurately select the appropriate species during Space Ranger execution. For GeoMx, the GeoMx NGS Pipeline developed by NanoString is required to convert FASTQ to digital count conversion (DCC) files, which can be further processed by SpaceSquest. In the case of imaging-based technologies such as Xenium and CosMx, files generated by the platform can be readily used for SpaceSquest data processing.

2.2 CLUSTERING

Clustering is a powerful tool for understanding the structure of the data, as well as the spatial tissue domains. For general clustering of single cell expression in CosMx and Xenium, or bins in Visium HD, we utilized the UMAP dimension reduction followed by the FindNeighbors and FindClusters functions. For Visium technology, we have incorporated two more clustering tools, SpaGCN⁵² and BayesSpace⁵³. SpaGCN leverages graph convolutional networks to integrate gene expression, spatial location, and histology data, allowing for the identification of spatial domains with distinct expression patterns⁵². In contrast, BayesSpace is a Bayesian statistical tool that leverages prior spatial knowledge to cluster transcriptomically similar spots in proximity⁵³. Additionally, BayesSpace offers the capability to enhance resolution by reallocating gene counts from larger spots in array-based data, such as Visium, to more refined sub-spots based on spatial information from nearby spots⁵³.

2.3 CELL TYPE ANNOTATION

Automated cell type annotation can be achieved through transferring labels from a reference single-cell RNA-seq dataset. For single-cell-level assays, including CosMx and Xenium, we incorporated Azimuth to predict the most likely cell type for each cell⁵⁴. A UMAP plot will be generated after running the pipeline for users to examine the cell type annotation results.

2.4 DECONVOLUTION

For deconvolution purposes, we have incorporated two state-of-the-art deconvolution tools, Tangram⁵⁵ and Cell2location⁴². We selected these two packages for their outstanding performance in a recent benchmarking study⁴³. Tangram, a deep-learning framework, excelled in creating comprehensive single-cell-resolution spatial gene-expression maps and integrating histological and anatomical data from the same specimens. It adeptly handled low-quality genes in high-resolution methods, offered single-cell resolution for lower-resolution techniques, and ensured broad genome-wide coverage for targeted approaches⁵⁵. Meanwhile, Cell2location, a Bayesian model, showcased its ability to discern fine-grained cell types within spatial transcriptomic data and constructs detailed cellular maps of diverse tissues. In the first step, Cell2location estimated reference cell type signatures from single-cell and single-nucleus RNA-seq profiles, aligned with gene expression profiles for user-specified cell types and subpopulations. In the subsequent stage, Cell2location utilized these reference signatures in conjunction with one or more spatial transcriptomic datasets to deconstruct mRNA counts at individual spatial locations into their respective reference cell types⁴².

2.5 LIGAND RECEPTOR INTERACTIONS

Here, we employ SpaTalk to investigate spatially resolved cell-cell interactions that will enable us to get insights into tissue homeostasis, development, and disease⁵⁶. This tool leverages a graph network along with a knowledge graph to model and evaluate the ligand-receptor-target signaling network operating between neighboring cells⁵⁶. SpaTalk serves as an optional step in the Visium data processing workflow, providing researchers with a powerful means to unravel the complex dynamics of inter-cellular interactions within tissues⁵⁶.

2.6 STANDARDIZED OUTPUT

SpaceSquest generates a series of output files that are ready for result query and further analysis. The R objects are saved as RData files, and the python anndata objects are saved as h5ad files. These files contain expression matrices, dimension reduction results, and intermediate data generated during pipeline processing.

3. RESULT AND DISCUSSION

3.1 HIGHLIGHTS OF SPACESEQUENT

The pipeline incorporates several key workflows to perform various analyses corresponding to each platform. Key strengths of this pipeline include:

- 1) Self-explainable configuration files that ensure easy set up.
- 2) Standardized and reproducible data processing powered by R and Python.
- 3) Customized tools are integrated in the pipeline for each spatial platform.
- 4) Comprehensive result generation, including R/Python objects and key figures.
- 5) Streamlined data visualization enabled by cellxgene VIP and Quickomics.

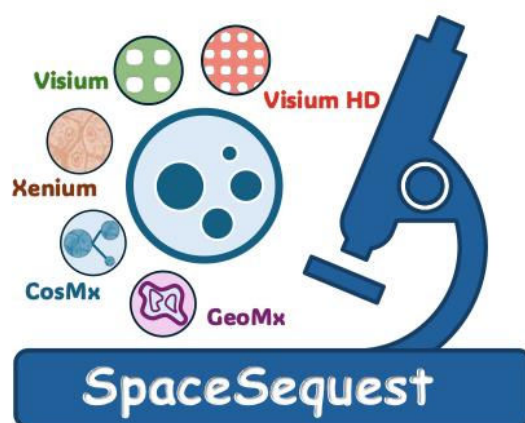


Fig. 1. Components of SpaceSequest. The pipeline integrates five major workflows tailored to perform a series of analysis for each platform, including Visium, Visium HD, Xenium, CosMx, and GeoMx.

3.2 OVERALL PIPELINE DESIGN

SpaceSequest is a unified toolkit which contains a set of workflows. Each workflow can be activated by calling the workflow script, and necessary configuration files will be created accordingly. The main configuration files are in YAML format, ensuring standardized formatting and flexible modification. To execute the pipeline, in most of the times, users are typically required to fill in a sample metadata file, which documents the details of each data, such as sample name, file path, and associated sample-level information. The overall pipeline components are illustrated in Figure 1.

3.3 VISIUM DATA ANALYSIS INCORPORATES MULTIPLE TOOLS

The 10x Genomics Visium platform was released in 2019 and continues to serve as a powerful commercial spatial transcriptomics technology since then. The Visium slides contain an array of barcoded 55 μm spots, with two major tissue capture sizes: 6.5x6.5 mm or 11x11 mm. Sequencing data in FASTQ format are processed through Space Ranger, and SpaceSequest continues the analysis by utilizing the binned output files. Firstly, SpaceSequest performs standard quality control on the data by removing spots with low gene detection rates, low UMI counts, etc. The filtered data are then integrated into a new h5ad file, which is piped to downstream tools for further analysis.

We tested our pipeline on a public dataset generated from the brain tissue of an Alzheimer's disease mouse model, which has mutations in APP and PSEN1 genes⁵⁷. We also downloaded the Azimuth mouse motor cortex single-nucleus RNA-seq data as a reference for cell type annotation⁵⁸. The data analysis is achieved by calling the *visium* command in SpaceSequest, which reads all input data and performs a series of quality control steps (Figure 2A). The cleaned data are then integrated for downstream processing, which includes spatial clustering and cell type deconvolution results. To leverage the power of parallel computing, SpaceSequest is able to submit sbatch jobs to execute the following four tools simultaneously: BayesSpace⁵³, Cell2location⁴², SpaGCN⁵², and Tangram⁵⁵. The jobs will be monitored until results have been generated and saved in separate directories. To provide a few examples, BayesSpace clustering step will cluster the spots across samples and identify comparable spatial clusters (Figure 2B). While Cell2location results can be visualized by cell abundance heatmaps overlayed with the tissue image (Figure 2C). Additionally, the pipeline utilizes SpaTalk for cell-cell communication analysis, which requires Cell2location output so it is only executed once Cell2location results are available (Figure S1)⁵⁶. Overall, our Visium data analysis workflow ensures automated data processing and analysis with a streamlined, one-stop-shop design (Figure 2A).

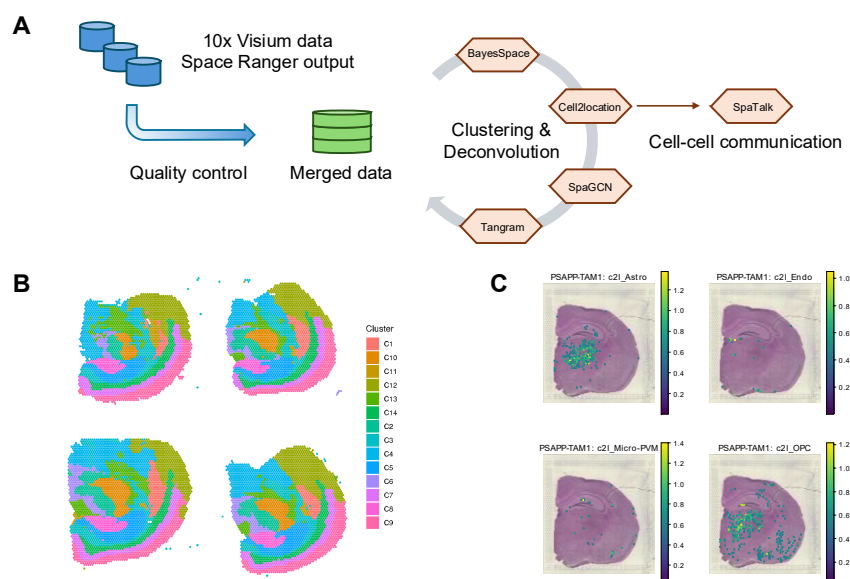


Fig. 2. 10x Visium data analysis. (A) Workflow for Visium data processing. (B) Spatial clustering results generated by BayesSpace. This run contains four samples, and BayesSpace identified shared clusters across them. (C) Predicted cell type abundance by Cell2location in one sample (PSAPP-TAM1), highlighting astrocytes, OPCs, microglia, and endothelial cells.

3.4 STREAMLINED XENIUM AND COSMX DATA ANALYSIS

The 10x Genomics Xenium and NanoString (now acquired by the Bruker Corporation) CosMx platforms are both imaging-based, high-plex in-situ technologies capable of capturing up to thousands of RNAs simultaneously at single-cell resolution. They require a set of probes, usually from pre-designed panels, with additional space for customization as well. Xenium data consists of a set of well-defined files in various formats, such as the `cell_feature_matrix` files (a folder with three associated files, or a single h5 file), transcript files (`transcripts.csv.gz`, `transcripts.parquet`, `transcripts.zarr.zip`), and cell summary files (`cells.csv.gz`, `cells.parquet`), etc. The `LoadXenium` function from the Seurat package reads the necessary files from the directory and converts them into a Seurat object for downstream analysis.

CosMx data are stored as a collection of plain text files in the comma-separated values (csv) format. These files include a metadata file (`metadata_file`), a file storing field of view (fov) positions (`fov_positions_file`), an expression matrix file (`exprMat_file`), a polygon file (`polygons`), and a transcript data file (`tx_file`). Data loading is achieved by the `LoadNanostring` function, which converts these csv files to a Seurat object.

SpaceSequest contains two workflows—*xenium* and *cosmx*—to streamline the process from data loading to downstream analysis. Once Xenium or CosMx data are successfully loaded, these workflows incorporate standard procedures that perform quality control, clustering, cell type annotation, and differential expression (DE) analysis, etc (Figure 3A). Moreover, the workflows generate UMAP and tissue visualizations, enabling users to examine the dimension reduction and cell type annotation results. Example plots generated using public datasets are shown in Figure 3B-D. The Xenium data were generated from a healthy human brain FFPE sample profiled with the Xenium Human Brain Gene Expression Panel containing 266 targets. On the other hand, the CosMx data were derived from human frontal cortex FFPE tissue using >6,000 target genes. After running the workflow, various analysis results and the processed Seurat object will be saved for future queries and additional analyses, including further exploration by Cellxgene VIP⁵⁹.

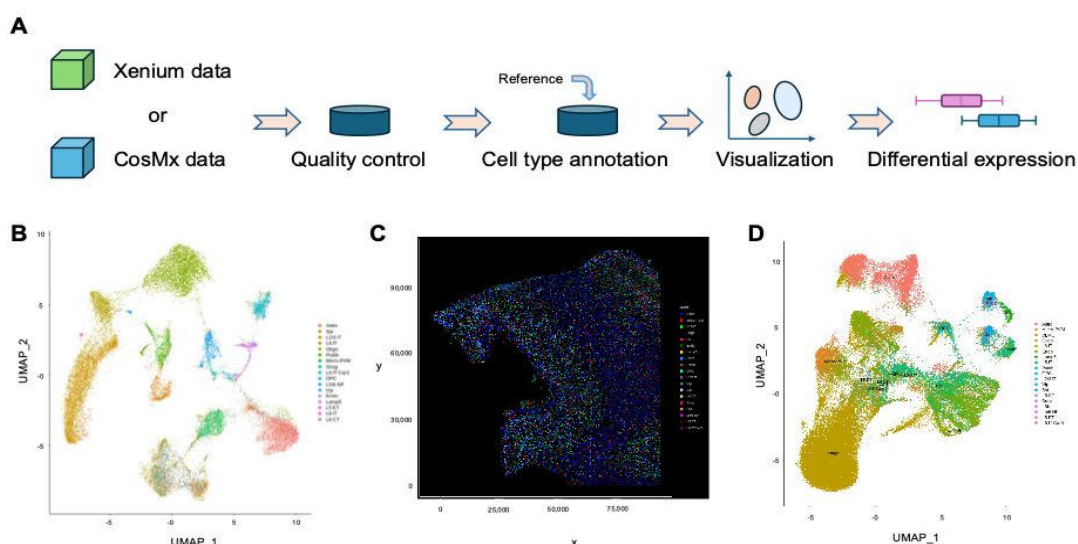


Fig. 3. Xenium and CosMx data analysis. (A) Streamlined workflows for analyzing Xenium and CosMx data. (B) UMAP visualization of Xenium data, labelled by predicted cell types. (C) Tissue and cell visualization of CosMx data. (D) UMAP visualization of CosMx data with predicted cell types labelled.

3.5 ENHANCED GEOMX DATA PROCESSING AND VISUALIZATION

The NanoString GeoMx platform is a useful technique to compare gene expression differences in defined regions on the tissue slides. It uses probes to detect the whole transcriptome in the selected region of interests (ROIs). However, GeoMx is not a single-cell resolution technique; instead, it requires a sufficient number of cells within the region to generate robust signals. Thus, GeoMx is considered as a mini-bulk RNA detection platform, capable of profiling >18,000 major genes for human and mouse samples. In our pipeline, we developed *geomx*, a workflow that takes DCC files, sample metadata, and a set of comparison groups as input. It performs a series of analysis, and outputs differential expression analysis results (Figure 4A). This workflow generates four csv files that are fully compatible with other tools including some Shiny applications^{60,61}, for example, Quickomics, an open-source R Shiny application for flexible data exploration and visualization⁶⁰. After uploading the data, users can examine PCA plots (Figure 4B), volcano plots (Figure 4C), heatmaps (Figure 4D), and a number of on-the-fly visualizations through Quickomics. This strategy not only automates GeoMx data analysis, especially for studies with many samples, but bridges the gap between raw data and insight generation through Quickomics, which creates a link to facilitate data storage and future analysis.

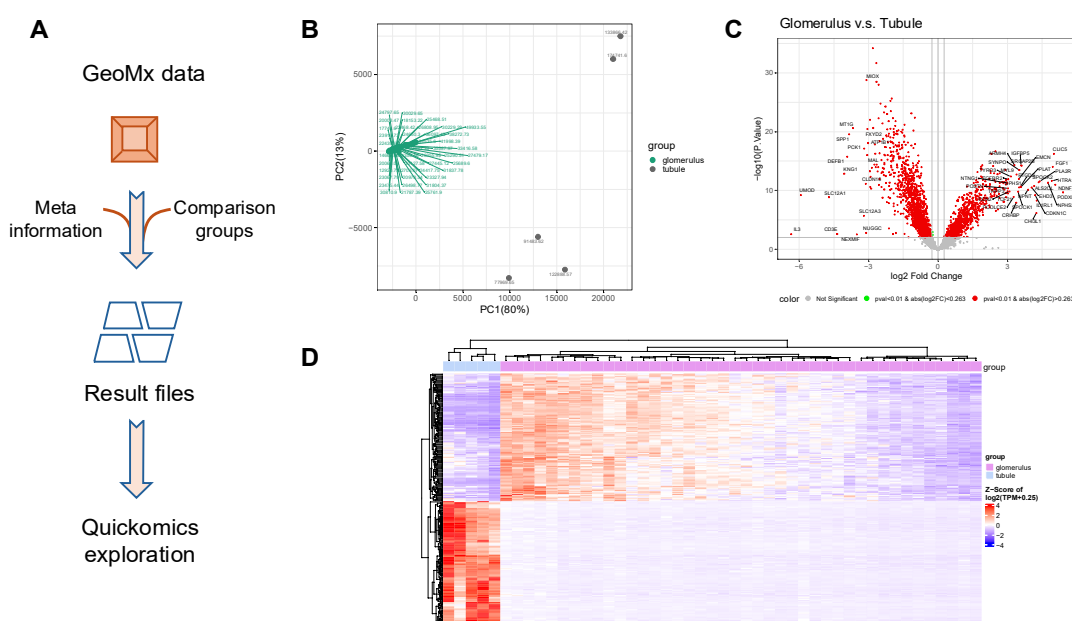


Fig. 4. GeoMx data analysis. (A) Overview of GeoMx data analysis workflow. (B) PCA plot showing samples from two kidney regions, glomerulus and tubule. (C) Volcano plot highlighting differentially expressed genes between glomerulus

and tubule. (D) Heatmap displaying the top differentially expressed genes between glomerulus and tubule, with samples and genes clustered by dendrograms.

3.6 STANDARDIZED DATA ANALYSIS OF THE CUTTING-EDGE VISIUM HD PLATFORM

In 2024, 10x Genomics released a new platform called Visium HD, featuring high-resolution, gap-free detection of tissue RNAs. This platform is now compatible with FFPE, fixed frozen, and fresh frozen samples. To keep pace with the rapid evolution of spatial transcriptomics technologies, we incorporated *visiumhd* into SpaceSequest for Visium HD data processing. The workflow takes Space Ranger outputs, performs bin-level quality control, dimension reduction, and cell type annotation when an Azimuth reference dataset is available (Figure 5A). By executing the pipeline, users receive plots to examine the gene counts (Figure 5B) and cell type annotation results (Figure 5C). Furthermore, we developed a new function, *getVisiumHDspotCol*, to extract the mean RGB intensities when a paired high-resolution immunofluorescence image is available (Figure 5D). This function facilitates accurate bin selection using nuclei staining, cell type marker staining, or other biomarker staining, for example, pathological protein markers.

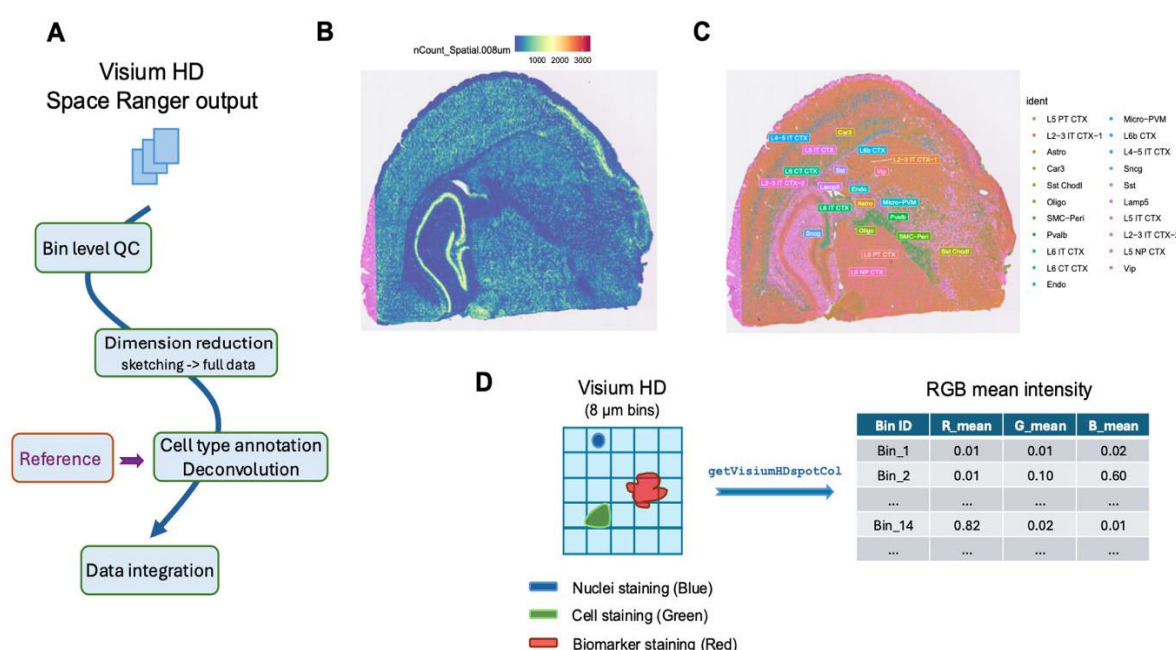


Fig. 5. Visium HD data analysis. (A) Overview of the general processing workflow for Visium HD data. (B) Heatmap showing the number of UMI counts (nCount) detected across the tissue section at 8 μ m bin resolution. (C) Visualization of annotated cell types on the tissue section. (D) Illustration of a new function that calculates the average RGB intensities across bins.

4. CONCLUSION

In this study, we developed SpaceSequest, a unified pipeline for spatial transcriptomics data processing. Currently, it supports five major platforms and incorporates customized software and packages to maximize the information that can be extracted from the data. To make the pipeline accessible, we provided a detailed tutorial aimed at lowering the barrier for researchers to apply this SpaceSequest to their own studies. Guided by the FAIR principles (Findable, Accessible, Interoperable, and Reusable), SpaceSequest enhances data access, interpretation, and facilitates data sharing by generating standardized results. In summary, SpaceSequest is a user-friendly, end-to-end solution for spatial transcriptomics data analysis and interpretation.

AVAILABILITY OF DATA AND MATERIALS

No new data were generated from this study. Several public datasets were used to demonstrate SpaceSequest. The 10x Visium data were from a public study using mouse brain tissues with NCBI accession number GSE203424 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE203424>). Public data were downloaded from 10x Genomics dataset webpage (<https://www.10xgenomics.com/datasets>) to test Xenium (<https://www.10xgenomics.com/datasets/xenium-human-brain-preview-data-1-standard>) and Visium HD (FFPE: <https://www.10xgenomics.com/datasets/visium-hd-cytassist-gene-expression-libraries-of-mouse-brain-he>, Fixed frozen: <https://www.10xgenomics.com/datasets/visium-hd-cytassist-gene-expression-mouse-brain-fixed-frozen>) workflows,

© Author(s) 2026. This work is distributed under the Creative Commons BY- 4.0 license:

<https://creativecommons.org/licenses/by/4.0/>

respectively. Moreover, we used a public human kidney GeoMx dataset (http://nanosttring-public-share.s3-website-uswest-2.amazonaws.com/GeoScriptHub/Kidney_Dataset_for_GeomxTools.zip) and a human brain CosMx dataset (<https://nanosttring.com/products/cosmx-spatial-molecular-imager/ffpe-dataset/human-frontal-cortex-ffpe-dataset/>) from NanoString websites to test GeoMx and CosMx pipelines respectively. More details related to data downloading and processing can be found at the tutorial (<https://interactivereport.github.io/SpaceSequest/tutorial/docs/index.html>) associated with this pipeline.

COMPETING INTERESTS

Y.H.S., S.P., Y.C., J.G., S.C., B.S., J.Z., W.H., F.C., D.H., and B.Z. are employees of Biogen and hold stocks from the company. H.Z., K.C. and H.Y. were co-ops at Biogen and do not hold Biogen's stocks. Z.O. and M.R. are employees of BioInfoRx Inc. and PharmaLex, respectively.

ACKNOWLEDGEMENT

This study is funded by Biogen Inc. We acknowledge the scientists at Biogen for their valuable feedback on the pipeline.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Examples of SpaTalk results. (A) Highlighted pairs between astrocytes (Astro) and GABAergic neurons. (B) Significantly enriched ligand-receptor interaction (LRI) pairs between astrocytes and GABAergic neurons. (C) Highlighted pairs between microglia and macrophage cells (Micro_PVM) and Glutamatergic neurons. (D) Significantly enriched ligand-receptor interaction (LRI) pairs between microglia/macrophages and Glutamatergic neurons.

REFERENCES

- [1] Jovic, D. *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical & Translational Med* 12, e694 (2022).
- [2] Wang, S. *et al.* The Evolution of Single-Cell RNA Sequencing Technology and Application: Progress and Perspectives. *IJMS* 24, 2943 (2023).
- [3] Potter, S. S. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol* 14, 479–492 (2018).
- [4] Du, J. *et al.* Advances in spatial transcriptomics and related data analysis strategies. *J Transl Med* 21, 330 (2023).
- [5] Adema, K., Schon, M. A., Nodine, M. D. & Kohlen, W. Lost in space: what single-cell RNA sequencing cannot tell you. *Trends in Plant Science* 29, 1018–1028 (2024).
- [6] Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R. & Haque, A. An introduction to spatial transcriptomics for biomedical research. *Genome Med* 14, 68 (2022).
- [7] Longo, S. K., Guo, M. G., Ji, A. L. & Khavari, P. A. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat Rev Genet* 22, 627–644 (2021).
- [8] Saviano, A., Henderson, N. C. & Baumert, T. F. Single-cell genomics and spatial transcriptomics: Discovery of novel cell states and cellular interactions in liver physiology and disease biology. *Journal of Hepatology* 73, 1219–1230 (2020).
- [9] Li, X. & Wang, C.-Y. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci* 13, 36 (2021).
- [10] Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat Methods* 19, 534–546 (2022).
- [11] Tian, L., Chen, F. & Macosko, E. Z. The expanding vistas of spatial transcriptomics. *Nat Biotechnol* 41, 773–782 (2023).
- [12] Wang, Y. *et al.* Spatial transcriptomics: Technologies, applications and experimental considerations. *Genomics* 115, 110671 (2023).
- [13] Jain, S. & Eadon, M. T. Spatial transcriptomics in health and disease. *Nat Rev Nephrol* 20, 659–671 (2024).
- [14] Piwecka, M., Rajewsky, N. & Rybak-Wolf, A. Single-cell and spatial transcriptomics: deciphering brain complexity in health and disease. *Nat Rev Neurol* 19, 346–362 (2023).
- [15] Zhang, L., Xiong, Z. & Xiao, M. A Review of the Application of Spatial Transcriptomics in Neuroscience. *Interdiscip Sci Comput Life Sci* 16, 243–260 (2024).
- [16] Jung, N. & Kim, T.-K. Spatial transcriptomics in neuroscience. *Exp Mol Med* 55, 2105–2115 (2023).
- [17] Tsai, H.-H. *et al.* Spatial transcriptomics reveals heterogeneous cell–cell interactions among brain regions in a cuprizone model consistent with multiple sclerosis lesions. *bioRxiv* 2025.02.07.637125 (2025) doi:10.1101/2025.02.07.637125.
- [18] Berglund, E. *et al.* Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun* 9, 2419 (2018).
- [19] Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* 24, 494–515 (2023).
- [20] Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat Biotechnol* 40, 1349–1359 (2022).
- [21] Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution. *Nature* 622, 552–561 (2023).

- [22] Huuki-Myers, L. A. *et al.* A data-driven single-cell and spatial transcriptomic map of the human prefrontal cortex. *Science* 384, eadh1938 (2024).
- [23] Qian, X. *et al.* Spatial transcriptomics reveals human cortical layer and area specification. *Nature* <https://doi.org/10.1038/s41586-025-09010-1> (2025) doi:10.1038/s41586-025-09010-1.
- [24] Han, L. *et al.* Single-cell spatial transcriptomic atlas of the whole mouse brain. *Neuron* 113, 2141–2160.e9 (2025).
- [25] Lake, B. B. *et al.* An atlas of healthy and injured cell states and niches in the human kidney. *Nature* 619, 585–594 (2023).
- [26] Wu, H. *et al.* Integrating spatial transcriptomics with single-cell transcriptomics reveals a spatiotemporal gene landscape of the human developing kidney. *Cell Biosci* 12, 80 (2022).
- [27] Chen, D. *et al.* Single-cell RNA -seq with spatial transcriptomics to create an atlas of human diabetic kidney disease. *The FASEB Journal* 37, e22938 (2023).
- [28] Mohenska, M. *et al.* 3D-cardiomics: A spatial transcriptional atlas of the mammalian heart. *Journal of Molecular and Cellular Cardiology* 163, 20–32 (2022).
- [29] Lee, S. E. *et al.* Spatial transcriptional landscape of human heart failure. *European Heart Journal* ehaf272 (2025) doi:10.1093/eurheartj/ehaf272.
- [30] Chen, H. *et al.* Dissecting mammalian spermatogenesis using spatial transcriptomics. *Cell Reports* 37, 109915 (2021).
- [31] Madissoon, E. *et al.* A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. *Nat Genet* 55, 66–77 (2023).
- [32] Vannan, A. *et al.* Spatial transcriptomics identifies molecular niche dysregulation associated with distal lung remodeling in pulmonary fibrosis. *Nat Genet* 57, 647–658 (2025).
- [33] Zhang, Y. *et al.* Reference-based cell type matching of in situ image-based spatial transcriptomics data on primary visual cortex of mouse brain. *Sci Rep* 13, 9567 (2023).
- [34] Chen, W.-T. *et al.* Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer’s Disease. *Cell* 182, 976–991.e19 (2020).
- [35] Yu, M. *et al.* Spatial transcriptomic patterns underlying amyloid- β and tau pathology are associated with cognitive dysfunction in Alzheimer’s disease. *Cell Reports* 43, 113691 (2024).
- [36] Navarro, J. F. *et al.* Spatial Transcriptomics Reveals Genes Associated with Dysregulated Mitochondrial Functions and Stress Signaling in Alzheimer Disease. *iScience* 23, 101556 (2020).
- [37] Zheng, P. *et al.* Integrated spatial transcriptome and metabolism study reveals metabolic heterogeneity in human injured brain. *Cell Reports Medicine* 4, 101057 (2023).
- [38] Ferri-Borgogno, S. *et al.* Spatial Transcriptomics Depict Ligand–Receptor Cross-talk Heterogeneity at the Tumor-Stroma Interface in Long-Term Ovarian Cancer Survivors. *Cancer Research* 83, 1503–1516 (2023).
- [39] Moreno, J. *et al.* Identification of ligand and receptor interactions in CKD and MASH through the integration of single cell and spatial transcriptomics. *PLoS ONE* 19, e0302853 (2024).
- [40] Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363, 1463–1467 (2019).
- [41] Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods* 16, 987–990 (2019).
- [42] Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* 40, 661–671 (2022).
- [43] Li, H. *et al.* A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nat Commun* 14, 1548 (2023).
- [44] Navarro, J. F., Sjöstrand, J., Salmén, F., Lundeberg, J. & Ståhl, P. L. ST Pipeline: an automated pipeline for spatial mapping of unique transcripts. *Bioinformatics* 33, 2591–2593 (2017).
- [45] Sztanka-Toth, T. R., Jens, M., Karaikos, N. & Rajewsky, N. Spacemake: processing and analysis of large-scale spatial transcriptomics data. *GigaScience* 11, giac064 (2022).
- [46] Moriel, N. *et al.* NovoSpaRc: flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nat Protoc* 16, 4177–4200 (2021).
- [47] Cisar, C., Keener, N., Ruffalo, M. & Paten, B. A unified pipeline for FISH spatial transcriptomics. *Cell Genomics* 3, 100384 (2023).
- [48] Kamel, M. *et al.* ENACT: End-to-End Analysis of Visium High Definition (HD) Data. *Bioinformatics* 41, btaf094 (2025).
- [49] Curion, F. *et al.* Panpipes: a pipeline for multiomic single-cell and spatial transcriptomic data analysis. *Genome Biol* 25, 181 (2024).
- [50] Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 42, 293–304 (2024).
- [51] Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Wolf, F. A. anndata: Annotated data. *bioRxiv* 2021.12.16.473007 (2021) doi:10.1101/2021.12.16.473007.
- [52] Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* 18, 1342–1351 (2021).

- [53] Zhao, E. *et al.* Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* 39, 1375–1384 (2021).
- [54] Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29 (2021).
- [55] Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods* 18, 1352–1362 (2021).
- [56] Shao, X. *et al.* Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat Commun* 13, 4429 (2022).
- [57] Castranio, E. L. *et al.* Microglial *INPP5D* limits plaque formation and glial reactivity in the PSAPP mouse model of Alzheimer's disease. *Alzheimer's & Dementia* 19, 2239–2252 (2023).
- [58] Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* 184, 3222–3241.e26 (2021).
- [59] Li, K. *et al.* Cellxgene VIP unleashes full power of interactive visualization and integrative analysis of scRNA-seq, spatial transcriptomics, and multiome data. *bioRxiv* 2020.08.28.270652 (2022) doi:10.1101/2020.08.28.270652.
- [60] Gao, B. *et al.* Quickomics: exploring omics data in an intuitive, interactive and informative manner. *Bioinformatics* 37, 3670–3672 (2021).
- [61] Gao, B. *et al.* xOmicsShiny: an R Shiny application for cross-omics data analysis and pathway mapping. *Bioinformatics Advances* 5, vbaf097 (2024).