Nucleic Acids Research, 2026, **54**, D1281–D1290
https://doi.org/10.1093/nar/gkaf998
Advance access publication date: 8 October 2025
NAR Breakthrough Article

So3D: a comprehensive three-dimensional spatial omics resource for decoding tissue architecture in physiology and disease

Hongying Zhao*,†, Xiangzhe Yin†, Siyao Wang, Zhichao Geng, Wentong Yu, Hongzheng Yu, Shangwei Ning *, Li Wang *

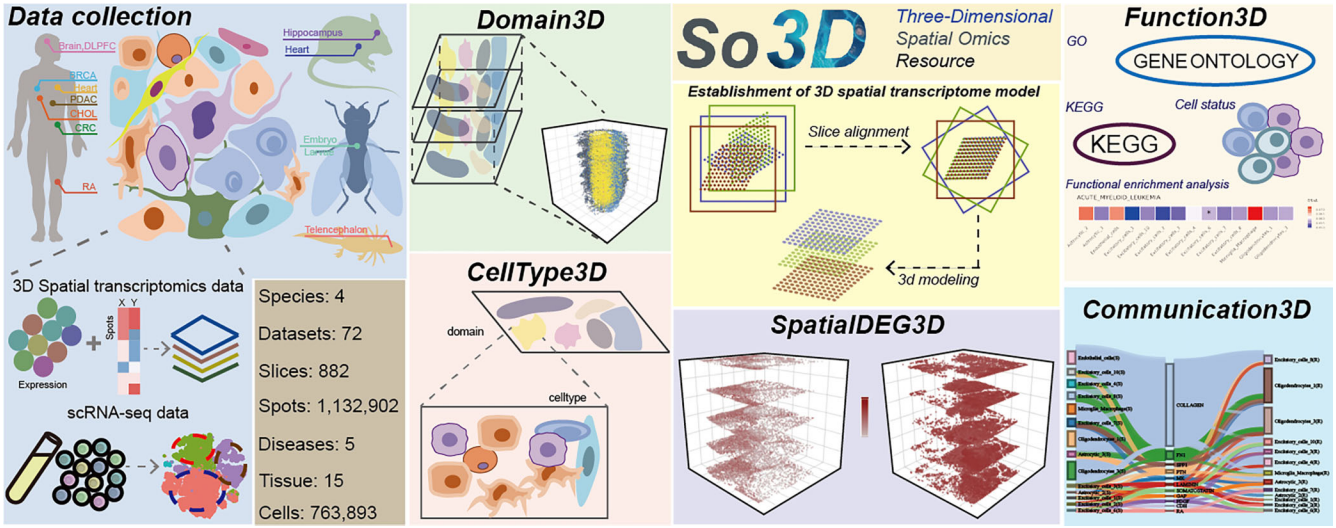
College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China

*To whom correspondence should be addressed. Email: wangli@hrbmu.edu.cn
Correspondence may also be addressed to Shangwei Ning. Email: ningsw@ems.hrbmu.edu.cn.
Correspondence may also be addressed to Hongying Zhao. Email: zhaohongying@hrbmu.edu.cn
†The first two authors should be regarded as Joint First Authors.

Abstract

Cells function within intricate three-dimensional (3D) architectures to form tissues and organs. Constructing 3D tissue structure is critical for understanding cellular states, biological processes, and intercellular interactions. Herein, we describe a comprehensive 3D spatial omics resource, So3D (http://bio-bigdata.hrbmu.edu.cn/So3D or https://so3d.bio-database.com/), which aims to construct 3D tissue architecture and dissect biological processes occurring in 3D space from multiple perspectives. We systematically collected 72 sets of 3D spatial transcriptome datasets, involving 1 132 902 spots across 882 slices from four species, and matched reference single-cell RNA-sequencing data, covering 763 893 cells from 28 datasets. So3D also provides multiple flexible analysis modules for retrieving and analyzing 3D tissue, such as inference of 3D spatial domains and gene expression patterns across 3D tissue regions, 2D spatial slices, and single-cell datasets; mapping of cell type distribution and cell–cell communication networks to explore intercellular crosstalk in 3D space; and functional annotation of biological pathways and signaling networks within 3D tissue contexts. Collectively, So3D provides comprehensive insights for investigating biologically meaningful 3D spatial domains, mapping local gene expression landscapes, functional state, and communication networks in tissue, and may also serve as an efficient and reliable tool for understanding the tissue microenvironment and discovering biomarkers.

Graphical abstract



Introduction

Three-dimensional (3D) spatial omics is becoming the "next-generation microscope" for understanding the life complexity, conquering diseases, and designing biological systems, and its application boundary is still expanding rapidly [1, 2]. By integrating molecular information with precise 3D position in

Received: August 12, 2025. **Revised:** September 7, 2025. **Accepted:** September 10, 2025

tissue or organ, 3D spatial omics reveals the biological laws of the "spatial dimension" that cannot be captured by conventional omics. Notably, the 3D spatial transcriptomics technology has constructed the 3D view of gene expression, cell types, and functional states within tissues and organs, revealing spatially continuous structures and biomarkers that are invisible in two-dimensional (2D) slices [3]. For example, 3D tissue reconstruction and structural analysis of breast cancer liver metastases have revealed independent subvolumes, demonstrating spatially heterogeneous growth patterns, invasive trajectories, and complex branching morphologies of the tumor. Furthermore, specific molecular signatures have been identified across different regions as follows: the proangiogenic factor TYMP and the gene IGLC2, which is related to immunoglobulin receptor binding [4]. Leveraging 3D models, researchers have further identified the spatial cell states of *Drosophila* embryos and predicted potential transcription factor regulators [5].

To elucidate the spatial regulation of gene expression and decode the spatial organization of cells, several studies have been conducted using spatial transcriptomics data. Current spatial transcriptomics databases such as STOmicsDB [6], SORC [7], CROST [8], SpatialDB [9], SpatialRef [10], and SPASCER [11] primarily focus on collecting spatial transcriptomic datasets across diverse species, organs, and diseases, and encompass a standardized spatial transcriptome data processing pipeline. However, biological organs inherently exhibit complex 3D architectures, and most biological processes occur in the 3D space rather than on a single 2D slice. These databases are limited to analyzing individual 2D slices and lack the capability to construct or analyze 3D tissue architectures. Currently, there is no integrated 3D spatial omics database that combines data collection, visualization, and systematic functional analysis. Existing research only provides fragmented online access to 3D spatial transcriptomics results. For example, VT3D is a visualization toolbox that enables 3D transcriptomic data browsing via surface model plots and has been applied to five published datasets (primarily *Drosophila*) [12]. Flysta3D allows users to query 3D spatial transcriptomics data in *Drosophila*, spanning from the embryonic to the pupal stages [5, 13]. 3DSeTH allows the visualization of gene expression using spatial transcriptomics data from six patients with rheumatoid arthritis [3]. Therefore, it is particularly urgent to develop a comprehensive resource for 3D spatial transcriptomics data collection, visualization, and systematic analysis.

In light of the above, we develop So3D, the first comprehensive, cross-species 3D spatial omics resource, which combines large-scale data integration with interactive 3D visualization and analysis tools to decode the spatial and functional complexity of biological systems across various physiological and disease states. So3D curates 72 sets of 3D spatial transcriptomics datasets, which contain 4 species, 15 tissues, 5 diseases, 882 slices, and 1 132 902 spots, alongside with matched the single-cell datasets, including 763 893 cells from 28 datasets. In more detail, So3D contains: (i) 30 human datasets from 10 different organs, covering five disease categories; (ii) 33 datasets from *Drosophila* embryos and larval tissues; (iii) 5 mouse datasets focused on brain, hippocampus, and heart tissues; and (iv) 4 datasets capturing the telencephalon tissue of axolotl. Furthermore, So3D offers five powerful analysis modules to gain in-depth biological exploration. Through an intuitive interactive interface, researchers can effortlessly perform data queries, visualization, and result downloads, thereby facilitating efficient study of 3D spatial transcriptomics data. This resource is invaluable for investigating 3D spatial domains, characterizing local gene expression, analyzing cell type interactions, and performing functional enrichment within tissues. Moreover, So3D supports key research applications including microenvironment analysis, biomarker discovery, and tissue biology studies, helping uncover regulatory mechanisms in tissues under physiological and pathological conditions. Collectively, So3D could serve as a pivotal resource for investigating 3D tissue architecture and mechanistic insights across diverse biological processes, such as embryogenesis, organogenesis, and disease progression, from a uniquely 3D vision.

Materials and methods

Three-dimensional spatial transcriptomics data collection

To systematically collect 3D spatial transcriptomics datasets, we performed a comprehensive search across multiple public repositories, comprising Gene Expression Omnibus, Human Tumor Atlas Network (https://data.humantumoratlas.org/), Mendeley Data (https://data.mendeley.com/), Zenodo (https://zenodo.org/), Spatial Transcript Omics DataBase (https://db.cngb.org/stomics/), and Single Cell PORTAL (https://singlecell.broadinstitute.org/), using keywords such as "spatial transcriptomics," "three-dimensional," and "3D." Additionally, we further screened recently published studies related to 3D spatial transcriptome analysis to ensure coverage of all available datasets based on consecutive sections. Ultimately, we collected 72 sets of 3D spatial transcriptomics datasets, spanning 4 species, 15 tissues, 5 diseases, 882 slices, and 1 132 902 spots.

Three-dimensional spatial transcriptomics data processing and clustering

For each 3D spatial transcriptomics dataset, we performed systematic data processing to construct the training model using the Python package STitch3D [14]. Specifically, we first incorporated the iterative closest point algorithm or PASTE algorithms to align multiple slices by spatial spot registration, thereby building the global 3D neighborhood map for spots in all slices. Then, we combined ST and single-cell RNA sequencing (scRNA-seq) data to eliminate the batch effect between slices. The model training was carried out using Adamax for stochastic optimization, employing the default parameters (20 000 optimization steps, learning rate $\mathrm{lr} = 0.002$, coefficients for computing running averages ($\beta_1 = 0.9$, $\beta_2 = 0.999$), and weight decay parameter $\lambda = 0$). After training, STitch3D output latent representations to perform 3D spatial domain detection. Next, we applied the Louvain algorithm to obtain the clustering result as spatial domains. For the Louvain algorithm, we tested Louvain in different resolutions using the Adjusted Rand Index (ARI) and selected the parameter resolution at 0.5 in the comparison (Supplementary Fig. S1). And this value has been widely adopted in spatial transcriptomics analyses to perform clustering analysis and is consistent with previous studies [15, 16]. Finally, we visualized the results using UMAP dimensionality reduction.

Identification of cell types

For each 3D spatial transcriptomics dataset, we conducted a joint analysis of spatially resolved gene expression matrices with 2D spatial coordinates from multiple consecutive ST slices and cell type–specific gene expression profiles from the paired scRNA-seq reference dataset. Leveraging the STitch3D deconvolution neural network, we accomplished 3D cell type deconvolution tasks by inferring cell type composition proportions at each spatial spot. Finally, the dominant cell type for each spot was defined based on the highest proportion.

scRNA-seq data collection and processing

For scRNA-seq data, we prioritized the use of scRNA-seq data matched with homologous samples for combined analysis. For spatial transcriptome datasets lacking paired scRNA-seq data, we selected the optimal scRNA-seq dataset, which has been used in previous studies for joint analysis with this spatial transcriptomics dataset. For scRNA-seq data processing, we employed the Seurat R package following standard quality control procedures [17]. The following criteria were then applied to select cells: (i) the number of expressed genes was >500; (ii) unique molecular identifier (UMI) count was >1000. Cells with low total UMI counts or few detected genes were filtered out. After raw data quality control and normalization, we also identified highly variable genes and performed principal component analysis for dimensionality reduction. We performed batch correction on the cells from different samples using the R package harmony [18], and applied the Louvain algorithm for clustering. UMAP visualization was used to characterize cell populations. Next, we manually curated tissue-specific cell type marker genes from the Annotation of Cell Types database [19]. We used the ScType algorithm to classify major cell populations based on these markers or define cell types based on existing biological knowledge [20]. For cancer samples specifically, we established stringent epithelial cell filtration criteria and applied the Cancer-Finder algorithm to accurately distinguish tumor cells from epithelial cells [21]. Marker genes of cell types were identified using the "FindAllMarkers" function in Seurat (min.pct > 0.25, $\log_2$(fold change) > 0.25, and padj < 0.05). A total of 763 893 cells from 28 datasets were integrated, covering 77 distinct cell types, thereby supporting high-precision cell type localization in 3D space.

Spatially differentially expressed genes

Detecting genes with spatially expression patterns in 3D space is critical for deciphering location-specific cellular states and functional mechanisms. First, we performed the "sc.pp.highly_variable_genes" function in Scanpy to identify highly variable genes. Next, we implemented the "sc.tl.rank_genes_groups" algorithm in Scanpy to systematically evaluate differential expression patterns across 3D spatial domains. Highly variable genes were defined as spatially differentially expressed genes (DEGs) if they met the following: pct_nz_group > 0.25, $\log_2$(fold change) > 0.25, and padj < 0.05.

Functional analysis

We performed Gene Ontology (GO) [22] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [23] enrichment analyses on cell type-specific DEGs of various species, including human, mouse, and *Drosophila*, using the clusterProfiler tool to uncover their underlying biological functions and pathway characteristics. To further investigate functional heterogeneity across cell types in 3D spatial transcriptomics, we computed single-sample gene set enrichment analysis scores for each spatial spot [24], incorporating GO and KEGG gene sets (human and mouse) from MSigDB database, and human cell state gene sets from CancerSEA [25]. Additionally, we applied hypergeometric testing to identify significantly enriched functions based on the DEGs of cell types (FDR < 0.05).

Communication analysis

To decipher cell–cell interaction networks among spatially adjacent cell types within 3D tissue architectures, we employed CellChat to systematically infer spatially proximal cell–cell communication from spatially resolved transcriptomic data [26]. We collected literature-supported ligand–receptor interactions from the built-in CellChatDB database, which includes both human and mouse data. Finally, we integrated aligned spatial coordinates and gene expression matrix from multiple tissue slices and ligand–receptor interaction database as input for analyzing communication patterns in 3D space.

Database framework and web implementation

So3D is freely available at http://bio-bigdata.hrbmu.edu.cn/So3D or https://So3D.bio-database.com/. The online web server of So3D was constructed by Java Server Pages language and deployed on Tomcat software (v9.0.93). All datasets of So3D were documented and managed based on the MySQL data source server (v5.6.50). To create the result data and realize the image visualization and interaction, several JavaScript packages were applied, including ECharts.min and ECharts-gl.min. All data processing and statistical analyses were performed using R software (v4.3.2) and Python software (v3.9). Currently, the So3D website can be well supported by state-of-the-art web browsers, such as Safari, Google Chrome, Microsoft Edge, and Firefox. No registration or login is required.

Database content and usage

Database overview

So3D is an integrated and user-friendly 3D spatial transcriptomics resource that constructs 3D tissue architectures and investigates biological processes occurring in 3D space from multiple perspectives. Through a systematic review of published papers up to 2025, So3D contains 72 high-quality 3D spatial transcriptome datasets, spanning 4 species (human, mouse, *Drosophila*, and axolotl), 15 tissue types, 5 disease categories, 882 tissue sections, and 1 132 902 spatial spots (Supplementary Table S1). To enable accurate cell type annotation for each spatial spot, So3D supplements with corresponding scRNA-seq references, covering 763 893 cells from 28 datasets (Fig. 1A and Supplementary Table S2). For each spatial transcriptome dataset, So3D provides a comprehensive description, such as species, organ, disease category, sequencing technology, number of tissue sections, and number of spots. Following quality control, the processed 3D spatial transcriptomics datasets contain an average of 15 734 spatial spots distributed across 12 tissue sections per dataset. Among

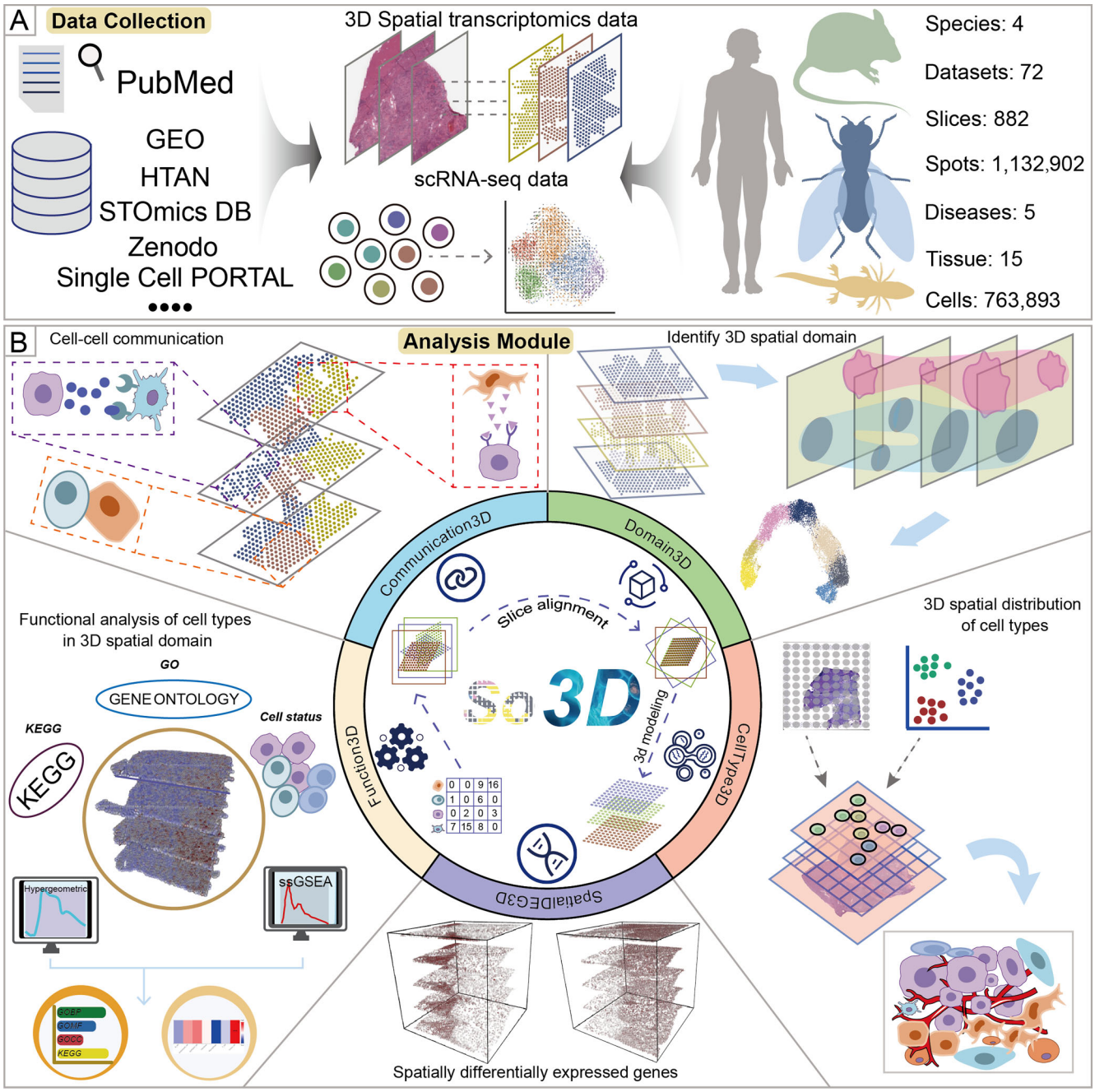


Figure 1. Overview of the design and statistics of So3D. (**A**) Data collection and construction of the So3D resource. (**B**) Analysis tools of the So3D resource.

them, the mouse heart dataset represents the most extensive section collection with 74 tissue sections, while the *Drosophila* larvae dataset contains the highest number of spatial spots at 63 603.

To elucidate the internal interactions and functional dynamics of complex biological systems, So3D introduces five tools to retrieve and analyze data: (i) Domain3D identifies 3D spatial domains with coherent gene expression patterns to annotate biologically meaningful tissue regions; (ii) CellType3D maps the 3D spatial distribution of cell types in tissues by integrating single-cell datasets, which provides an in-depth understanding of the possible biological functions in the enrichment regions of specific cell types; (iii) SpatialDEG3D provides spatially DEGs across 3D spatial domains and observes their expression levels in 2D sections and single-cell datasets, highlighting region-specific biomarkers; (iv) Function3D reveals biological functions and signaling pathways of cell type-specific enrichment, pinpointing the core function linked to the cell types; (v) Communication3D performs cell–cell communication to explore the crosstalk between cell types that are closely related in 3D space (Fig. 1B). Furthermore, So3D presents the results in the comprehensively intuitive chart format and offers an interactive 3D visualization module that supports multi-perspective data exploration, alongside seamless browsing, searching, and downloading of all available data.

User interface

Data browser

So3D offers efficient data browsing functions, enabling users to comprehensively explore the essential metadata of both 3D spatial transcriptome datasets and single-cell datasets, including species, disease categories, organ sources, sequencing techniques, slice counts, spot counts, cluster counts, and related literature. Users can also conduct multi-dimensional combination screening of datasets by species, disease category, organ sources, and sequencing techniques. Users can click on the target dataset to enter the dedicated analysis page and access detailed analysis results. In addition, So3D provides a global keyword search function to screen relevant data on a large scale (Fig. 2A).

Three-dimensional spatial domain identification

Annotating meaningful 3D spatial domains can accurately identify spatial functional units of biological significance. So3D provides biologically interpretable 3D spatial domains with similar gene expression profiles and spatial proximity. For Domain3D, users can view it intuitively as follows: (i) general information of dataset; (ii) 3D visualization and cross-section 2D comparative views of 3D spatial domains and the domain of interest; (iii) UMAP visualization based on 3D spatial domains and tissue sections; and (iv) specific expression of marker genes and statistical data in each 3D spatial domain. These functions provide a basis for detecting meaningful tissue structures related to biological functions and systematic investigation of regional functional heterogeneity (Fig. 2B).

Three-dimensional cell type spatial distribution

So3D provides precise cell type annotations for each spot in 3D spatial transcriptomics data, and further analyzes and reconstructs the 3D spatial distribution pattern of the cell types within tissues. By integrating the scRNA-seq reference data with spatial transcriptomics data, CellType3D offers the following: (i) interactive 3D visualization of the tissue distribution for all cell types and the cell type of interest; (ii) comparison view of cell type proportion distribution across 2D sections; (iii) display of the basic information, cell type annotation results, and cell type–specific markers of matched scRNA-seq data. CellType3D provides a systematic research tool for investigating the spatial heterogeneity in the tissue microenvironment and deciphering the enrichment regions of specific cell types (Fig. 2C).

Identification of spatially differentially expressed genes

Spatially DEGs serve as fundamental molecular determinants of 3D spatial heterogeneity. Identifying spatially DEGs is crucial for deciphering the functional specialization of distinct 3D spatial domains. Accordingly, SpatialDEG3D offers the following features: (i) a rigorously curated list of spatially DEGs and further statistics on the spatial domains and cell types where each gene is specifically enriched; (ii) visualization tools to compare their expression patterns across 3D space and corresponding 2D sections; and (iii) exploring expression levels of selected genes across different cell types in single-cell datasets. This module helps researchers precisely pinpoint region-specific biomarkers in 3D tissue regions and deeply reveal the functional differences among different regions (Fig. 2D).

Functional analysis of cell types in 3D space

To elucidate the functional roles of cell types within 3D spatial domain, So3D integrates four analytical modules for systematic characterization of cell type–specific functional enrichment: (i) Functional enrichment module for identifying significantly enriched biological functions and pathways among spatially DEGs across cell types; (ii) GO module for evaluating the activity and statistical significance with cell types of GO terms in 3D space, highlighting key biological processes associated with cell types; (iii) KEGG module for assessing the activity and statistical significance with cell types of KEGG pathway gene sets in 3D space; and (iv) Cell state module for detecting enrichment patterns of cell state gene sets in 3D space. Users can dynamically filter results by cell types or gene sets, presenting them in both tabular and graphical formats. Through Function3D, users enable precise identification of function linked to 3D cell types, facilitating mechanistic exploration (Fig. 2E).

Cell–cell communication in 3D space

Leveraging the comprehensive spatial relationship mapping capability of 3D spatial transcriptomics, we inferred the interactions among distinct cell types. Users can examine the significant interactions between cell types and the ligand–receptor pairs that mediate the interactions. All pairs are presented in the table, and the intensity and directionality of the interaction network are illustrated in the river diagram. Furthermore, Communication3D enables users to specify the ligand–receptor pair of interest and 3D visualization of the ligand and receptor expression and distribution in 3D space. Communication3D serves as a powerful tool for investigating the crosstalk between cell types in 3D space (Fig. 2F).

Example application of database

3D spatial analysis of breast cancer tissue

To systematically evaluate the potential application of So3D in deciphering 3D tissue architecture and uncovering novel biological insights, we analyzed 3D spatial transcriptomic data from breast cancer (BRCA) patients, covering 1729 spots across three consecutive 2D tissue sections (Fig. 3A). Using Domain3D, we identified five biologically coherent 3D spatial domains within the tumor tissue, defined as Cluster 0–4 (Fig. 3B). The consistency of these spatial domains across multiple sections demonstrated the accuracy of So3D in reconstructing 3D tissues across sections. Leveraging cell type signatures from scRNA-seq reference datasets, CellType3D deconvoluted the cellular composition of each spot, revealing distinct 3D spatial distributions of cell types (Fig. 3C). B cells and macrophages were specifically localized to Cluster 0, suggesting this domain as an immune-enriched region, while tumor cells dominated other spatial domains (including Cluster 1, Cluster 2, and Cluster 4), indicative of tumor-enriched regions. T cells were highly enriched in Cluster 3 (Fig. 3D). Subsequently, we used SpatialDEG3D to evaluate spatially DEGs in each 3D spatial region. Notably, the specific marker gene IGHA1 of B cells was significantly upregulated in Cluster 0 (Fig. 3E) [27]. The significantly upregulated gene DHCR24 in tumor cells was specifically highly expressed in Cluster 1

Figure 2. User interface of So3D database. (**A**) Data browser page allows users to obtain 3D spatial transcriptomics dataset based on species, disease type, tissue source, and sequencing technique. (**B**) Domain3D shows 3D spatial domains with coherent gene expression patterns and provides marker information in 3D spatial transcriptomics dataset. (**C**) CellType3D offers cellular component in 3D tissue and cell type annotation in single-cell dataset. (**D**) SpatialDEG3D offers spatially DEGs and shows the expression of the selected gene in 3D spatial domains and cell types. (**E**) Function3D provides cell type-specific functional enrichment. (**F**) Communication3D displays the significant interactions between cell types and the ligand–receptor pairs.

and Cluster 4 (Fig. 3E). This observation is consistent with the previous finding that DHCR24, a key enzyme in cholesterol biosynthesis, promotes breast cancer progression by enhancing tumor cell proliferation [28, 29]. Our tool successfully established strong correlations between 3D spatial cell type distributions and their corresponding marker genes, demonstrating its effectiveness in uncovering biologically meaningful spatial patterns. Analysis via Function3D module revealed elevated proliferative, cell cycle, and metastatic activity in tumor regions 1 and 4 of 3D breast cancer samples. Additionally, DHCR24-high 3D spatial regions 1 and 4 also exhibited significant enrichment of cholesterol biosynthesis and metabolism across consecutive tissue sections, further demonstrating the spatial regulatory role of DHCR24 in the pathogenesis of breast cancer (Fig. 3F). Finally, the Communication3D revealed that multiple ligand–receptor pairs mediate

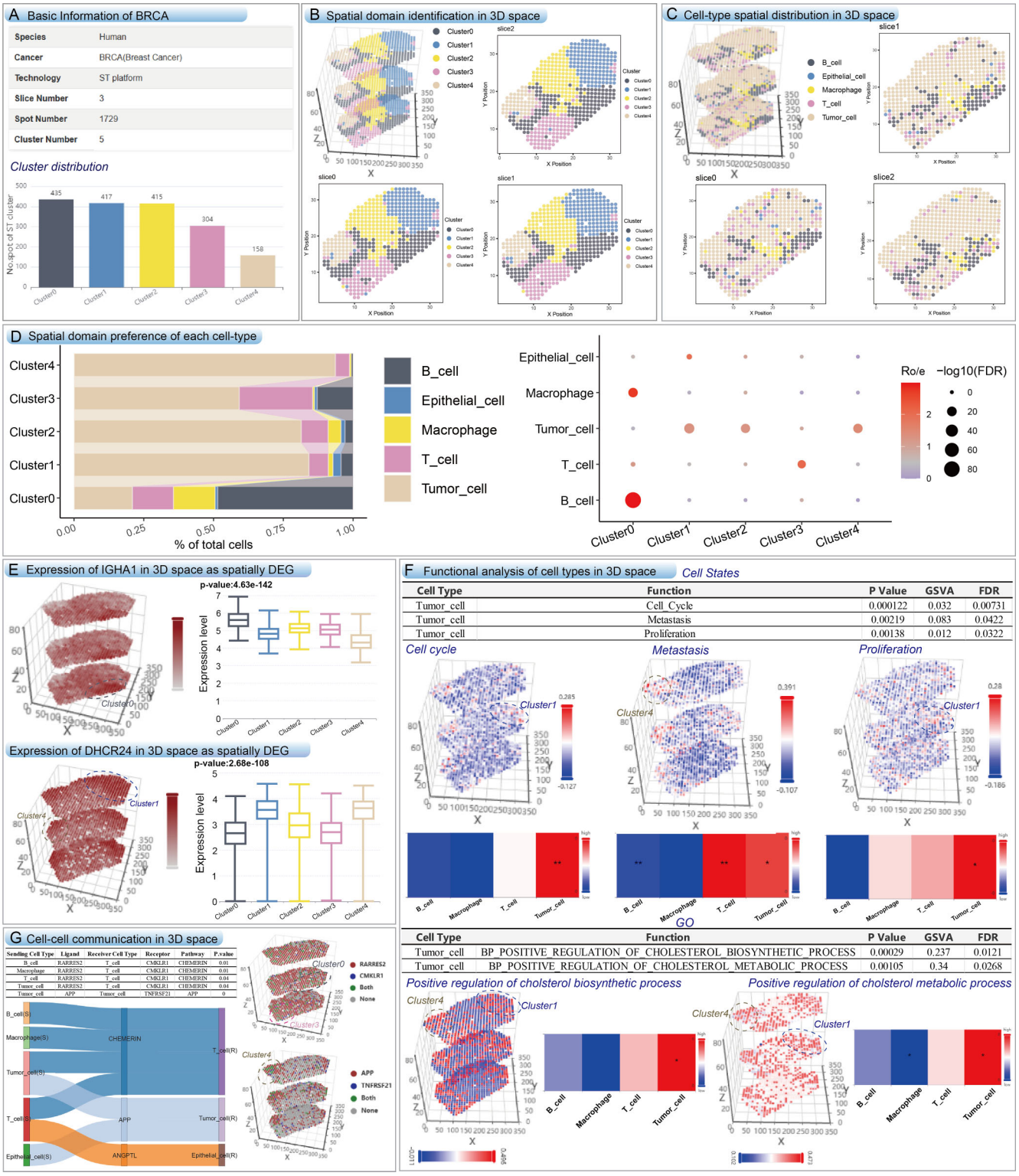


Figure 3. Case study from patients of the BRCA dataset in So3D. (**A**) The basic information and spot distribution of the BRCA dataset. (**B**) 3D spatial domain identification in 3D space and 2D slices. (**C**) 3D cell types distribution in 3D space and 2D slices. (**D**) 3D spatial domain distribution of each cell type. (**E**) The expression of IGHA1 and DHCR24 as spatially DEGs in 3D space. (**F**) Functional analysis results of tumor cells in 3D space. (**G**) The results of communication across cell types in 3D space and the expression distribution of ligand–receptor pairs.

the intercellular interactions in breast cancer tissues (Fig. 3G). For example, the interaction across immune cells was mediated by RARRES2–CMKLR1 pair to stimulate recruitment of immune cells and inhibit tumor growth, aligning with the chemerin signaling pathway reported previously [30]. In our analysis, the ligand RARRES2 and its receptor CMKLR1 were predominantly expressed in Cluster 0 (immune cell-enriched region) and Cluster 3 (a mixed region containing both immune and tumor cells), highlighting active immune communication within these areas. We also observed that tumor cells interact through the APP–TNFRSF21 axis, thereby promoting the formation of the immunosuppressive microenvironment to facilitate tumor progression [31–33]. Notably, APP and TNFRSF21 were simultaneously expressed in Cluster 4, a tumor cell-enriched region, suggesting an autocrine-like tumor–tumor signaling loop that may contribute to immune evasion.

3D transcriptomic landscape of the human dorsolateral prefrontal cortex

To further demonstrate the utility of So3D in constructing comprehensive 3D tissue panoramas, we analyzed the human dorsolateral prefrontal cortex dataset generated by the 10× Visium platform, comprising four tissue sections and 14 243 spots (Supplementary Fig. S2A). Using Domain3D, we identified seven spatial domains, which showed consistent patterns across all four slices and closely matched the manually annotated regions [34] (Supplementary Fig. S2B). Applying CellType3D, we visualized cell type deconvolution results and observed biologically meaningful distributions, in that oligodendrocyte subtype 1 was highly enriched in Cluster 2, whereas subtype 3 was enriched in Clusters 2 and 6, corresponding to white matter regions. Astrocyte subtypes were specifically enriched in Cluster 1, resembling the layer 1, while excitatory cells were widely distributed across Clusters 0, 1, 4, and 5, likely reflecting cortical layers 2–6 [34] (Supplementary Fig. S2C and D). Through SpatialDEG3D, we further visualized the expression of spatially DEGs. Marker genes of oligodendrocytes, such as PLP1 and MBP, were strongly expressed in Clusters 2, 3, and 6 (Supplementary Fig. S2E). Notably, MBP, a well-known white matter marker, displayed clear spatial restriction, further validating that So3D can effectively identify biologically relevant 3D spatial domains. To dissect the functional heterogeneity of cell types in 3D tissues, we applied Function3D, which revealed that oligodendrocyte development was significantly enriched in oligodendrocyte subtypes 1 and 3, and had a higher level of activity in the Cluster 2 (Supplementary Fig. S2F). Finally, Communication3D uncovered cell–cell interactions between oligodendrocytes and astrocytes mediated through collagen signaling (Supplementary Fig. S2G). Specifically, COL4A5 was highly expressed in oligodendrocyte-enriched regions (Cluster 2), while its receptors SDC4 and ITGAV were also significantly enriched in Cluster 2. These findings suggest that oligodendrocytes engage in astrocytic functions via the COL4A5 ligand–receptor pairs (SDC4, ITGB8, and ITGAV).

3D spatial transcriptomic atlas of the *Drosophila* embryo

Owing to rapid reproductive cycle, ease of genetic manipulation, and high genetic homology with humans, *Drosophila* serves as an ideal model for studying evolution, genetics, and disease. Here, we reconstructed the 3D embryo model of *Drosophila* in So3D, containing 21 tissue slices and 31 308 spots (Supplementary Fig. S3A). By integrating spatial transcriptomics data from these slices, Domain3D enabled the identification and visualization of seven consistent 3D spatial domains (Cluster 0–6) (Supplementary Fig. S3B). Leveraging scRNA-seq atlases, CellType3D generated the 3D spatial distribution of cell types within the embryo. We observed that the central nervous system (CNS) was specifically localized to Cluster 0 (Supplementary Fig. S3C and D), capturing the morphology and position of the embryonic CNS [13]. Both the foregut and the hindgut were mapped to Cluster 3 (Supplementary Fig. S3C and D), and their tissue localization during embryonic development is similar to that of the original studies. Interestingly, midgut-enriched regions (Clusters 4 and 5) were encapsulated within Cluster 3, reflecting their anatomical position between the foregut and the hindgut (Supplementary Fig. S3E and F). SpatialDEG3D further revealed biologically meaningful markers (Supplementary Fig. S3G). For example, Apolpp and Nplp2 were strongly expressed in the midgut-enriched regions (Cluster 4), highlighting the regulation of cellular metabolism and functions in lipid transport and storage [35–37]. In addition, we identified novel tissue marker genes, such as CG32073 and CG42586, broadly expressed in the midgut-enriched regions (Cluster 4). By Function3D, hindgut was found to be rich in protein synthesis-related functions, with ribosomal genes such as RpL38 and RpL41 expressed in both hindgut- and midgut-enriched regions (Clusters 2 and 4), pointing to potential cross-tissue functional connections (Supplementary Fig. S3H).

In summary, these findings validate So3D as a powerful tool for 3D research of biological systems to precisely localize key biological events within the microenvironment.

Discussion

3D tissue reconstruction marks a pivotal advance in spatial transcriptomics, providing unprecedented opportunities to study complex biological systems. While most existing databases are limited to 2D analyses [6–11], So3D represents the first comprehensive cross-species 3D spatial omics database that integrates large-scale datasets with interactive visualization and systematic analysis modules (Supplementary Table S3). The key innovation of So3D lies in its ability to accurately reconstruct tissue architecture from multi-slice datasets and to provide a unified platform that supports 3D spatial domain identification, precise cell type mapping, detection of spatially DEGs, functional annotation, and cell–cell communication analysis. These modules collectively enable researchers to decode the regulatory mechanisms governing tissue structure and function in the 3D context.

Recently, several benchmark methods, such as SpaMask [38], SpaDAMA [39], SpaDiT [40], mclSTExp [41], SpaICL [42], and SpaViT [43], have provided powerful computational strategies for advancing spatial transcriptomics analysis on individual 2D slice. So3D functions as a comprehensive resource that integrates large-scale, cross-species, multi-slice spatial transcriptomics datasets with multiple analytical modules and interactive visualization, thereby catering to this demand for reconstructing 3D panoramic views. Furthermore, So3D facilitates the exploration of tissue microenvironments, the identification of region-specific biomarkers, and the investigation of cellular interactions underlying the organ devel-

opment and disease progression. By offering intuitive visualization and efficient analytical workflows, So3D lowers the barrier for researchers to perform in-depth 3D spatial omics studies.

The rapid evolution of spatial multi-omics technologies (including metabolomics and genomics) has yielded exponentially growing high-resolution datasets. In the future, we will continue to update So3D, including (i) integrating the latest 3D spatial transcriptomic data; (ii) designing additional analysis modules to provide deep understanding of 3D tissues; and (iii) expansion of multi-omics data integration, such as spatial metabolomics and spatial genomics. In conclusion, by integrating multiple parallel 2D tissue sections, So3D achieves 3D tissue reconstruction, providing researchers with profound insights and new dimensions into tissue structure, local gene expression landscapes, cell interactions within the tissue, and biological functional regulation.

Acknowledgements

Author contributions: Hongying Zhao (Conceptualization [equal], Formal analysis [equal], Writing—original draft [equal], Writing—review & editing [equal]), Xiangzhe Yin (Conceptualization [equal], Formal analysis [equal], Methodology [equal], Validation [equal], Visualization [equal], Writing—original draft [equal]), Siyao Wang (Formal analysis [equal], Validation [equal], Visualization [equal]), Zhichao Geng (Formal analysis [equal], Validation [equal], Visualization [equal]), Wentong Yu (Formal analysis [equal], Validation [equal], Visualization [equal]), Hongzheng Yu (Formal analysis [equal], Validation [equal], Visualization [equal]), Shangwei Ning (Conceptualization [equal], Formal analysis [equal], Methodology [equal], Writing—review & editing [equal]), and Li Wang (Conceptualization [equal], Formal analysis [equal], Methodology [equal], Writing—review & editing [equal]).

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by National Natural Science Foundation of China (62372144, 62573169, 62572155) and Outstanding Youth Foundation of Heilongjiang Province of China (YQ2023F004).

Data availability

All the data used in the analysis can be obtained at http://bio-bigdata.hrbmu.edu.cn/So3D or https://so3d.bio-database.com/. So3D is freely accessible, without any registration requirements. The code for So3D is available at https://doi.org/10.5281/zenodo.16993653.

References

1. Schott M, Leon-Perinan D, Splendiani E *et al.* Open-ST: high-resolution spatial transcriptomics in 3D. *Cell* 2024;**187**:3953–72. https://doi.org/10.1016/j.cell.2024.05.055
2. Preibisch S, Innerberger M, Leon-Perinan D *et al.* Scalable image-based visualization and alignment of spatial transcriptomics datasets. *Cell Syst* 2025;**16**:101264. https://doi.org/10.1016/j.cels.2025.101264
3. Vickovic S, Schapiro D, Carlberg K *et al.* Three-dimensional spatial transcriptomics uncovers cell type localizations in the human rheumatoid arthritis synovium. *Commun Biol* 2022;**5**:129. https://doi.org/10.1038/s42003-022-03050-3
4. Mo CK, Liu J, Chen S *et al.* Tumour evolution and microenvironment interactions in 2D and 3D space. *Nature* 2024;**634**:1178–86. https://doi.org/10.1038/s41586-024-08087-4
5. Wang M, Hu Q, Lv T *et al.* High-resolution 3D spatiotemporal transcriptomic maps of developing *Drosophila* embryos and larvae. *Dev Cell* 2022;**57**:1271–83. https://doi.org/10.1016/j.devcel.2022.04.006
6. Xu Z, Wang W, Yang T *et al.* STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Res* 2024;**52**:D1053–61. https://doi.org/10.1093/nar/gkad933
7. Zhou W, Su M, Jiang T *et al.* SORC: an integrated spatial omics resource in cancer. *Nucleic Acids Res* 2024;**52**:D1429–37. https://doi.org/10.1093/nar/gkad820
8. Wang G, Wu S, Xiong Z *et al.* CROST: a comprehensive repository of spatial transcriptomics. *Nucleic Acids Res* 2024;**52**:D882–90. https://doi.org/10.1093/nar/gkad782
9. Fan Z, Chen R, Chen X. SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res* 2020;**48**:D233–7.
10. Cui T, Li YY, Li BL *et al.* SpatialRef: a reference of spatial omics with known spot annotation. *Nucleic Acids Res* 2025;**53**:D1215–23. https://doi.org/10.1093/nar/gkae892
11. Fan Z, Luo Y, Lu H *et al.* SPASCER: spatial transcriptomics annotation at single-cell resolution. *Nucleic Acids Res* 2023;**51**:D1138–49. https://doi.org/10.1093/nar/gkac889
12. Guo L, Li Y, Qi Y *et al.* VT3D: a visualization toolbox for 3D transcriptomic data. *J Genet Genomics* 2023;**50**:713–9. https://doi.org/10.1016/j.jgg.2023.04.001
13. Wang M, Hu Q, Tu Z *et al.* A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation. *Cell* 2025;**188**:4734–53. https://doi.org/10.1016/j.cell.2025.05.047
14. Wang G, Zhao J, Yan Y *et al.* Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. *Nat Mach Intell* 2023;**5**:1200–13. https://doi.org/10.1038/s42256-023-00734-1
15. Zhao J, Yu J, Cao Y *et al.* Enhancing spatial transcriptomics via spatially constrained matrix decomposition with EDGES. *Adv Sci* 2025;**21**:e08346. https://doi.org/10.1002/advs.202508346
16. Zhu J, Fan Y, Xiong Y *et al.* Delineating the dynamic evolution from preneoplasia to invasive lung adenocarcinoma by integrating single-cell RNA sequencing and spatial transcriptomics. *Exp Mol Med* 2022;**54**:2060–76. https://doi.org/10.1038/s12276-022-00896-9
17. Butler A, Hoffman P, Smibert P *et al.* Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411–20. https://doi.org/10.1038/nbt.4096
18. Korsunsky I, Millard N, Fan J *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019;**16**:1289–96. https://doi.org/10.1038/s41592-019-0619-0
19. Quan F, Liang X, Cheng M *et al.* Annotation of cell types (ACT): a convenient web server for cell type annotation. *Genome Med* 2023;**15**:91. https://doi.org/10.1186/s13073-023-01249-5
20. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from

single-cell transcriptomic data. *Nat Commun* 2022;**13**:1246. https://doi.org/10.1038/s41467-022-28803-w
21. Zhong Z, Hou J, Yao Z *et al.* Domain generalization enables general cancer cell annotation in single-cell and spatial transcriptomics. *Nat Commun* 2024;**15**:1929. https://doi.org/10.1038/s41467-024-46413-6
22. Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* 2015;**43**:D1049–56. https://doi.org/10.1093/nar/gku1179
23. Kanehisa M, Sato Y, Kawashima M *et al.* KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016;**44**:D457–62. https://doi.org/10.1093/nar/gkv1070
24. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;**14**:7. https://doi.org/10.1186/1471-2105-14-7
25. Yuan H, Yan M, Zhang G *et al.* CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res* 2019;**47**:D900–8. https://doi.org/10.1093/nar/gky939
26. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nat Protoc* 2025;**20**:180–219. https://doi.org/10.1038/s41596-024-01045-4
27. Hu C, Li T, Xu Y *et al.* CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res* 2023;**51**:D870–6. https://doi.org/10.1093/nar/gkac947
28. Yan L, Wu X, Zhang Y *et al.* LncRNA ENST00000370438 promotes cell proliferation by upregulating DHCR24 in breast cancer. *Mol Carcinog* 2023;**62**:855–65. https://doi.org/10.1002/mc.23529
29. Fu X, Wang Z. DHCR24 in tumor diagnosis and treatment: a comprehensive review. *Technol Cancer Res Treat* 2024;**23**:15330338241259780. https://doi.org/10.1177/15330338241259780
30. Rennier K, Shin WJ, Krug E *et al.* Chemerin reactivates PTEN and suppresses PD-L1 in tumor cells via modulation of a novel CMKLR1-mediated signaling cascade. *Clin Cancer Res* 2020;**26**:5019–35. https://doi.org/10.1158/1078-0432.CCR-19-4245
31. Zhao H, Yin X, Xu H *et al.* LncTarD 2.0: an updated comprehensive database for experimentally-supported functional lncRNA-target regulations in human diseases. *Nucleic Acids Res* 2023;**51**:D199–207. https://doi.org/10.1093/nar/gkac984
32. Zhao H, Shi J, Zhang Y *et al.* LncTarD: a manually-curated database of experimentally-supported functional lncRNA-target regulations in human diseases. *Nucleic Acids Res* 2020;**48**:D118–26.
33. He Y, Wang J, Deng Z *et al.* FOLR2^{+} macrophage depletion from intestinal metaplasia to early gastric cancer: single-cell sequencing insight into gastric cancer progression. *J Exp Clin Cancer Res* 2024;**43**:326. https://doi.org/10.1186/s13046-024-03245-y
34. Maynard KR, Collado-Torres L, Weber LM *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* 2021;**24**:425–36. https://doi.org/10.1038/s41593-020-00787-0
35. Rajan A, Karpac J. Inter-organ communication in *Drosophila*: lipoproteins, adipokines, and immune-metabolic coordination. *Curr Opin Cell Biol* 2025;**94**:102508. https://doi.org/10.1016/j.ceb.2025.102508
36. Rommelaere S, Boquete JP, Piton J *et al.* The exchangeable apolipoprotein Nplp2 sustains lipid flow and heat acclimation in *Drosophila*. *Cell Rep* 2019;**27**:886–99. https://doi.org/10.1016/j.celrep.2019.03.074
37. Niu C, Guan J, Meng G *et al.* The regulatory relationship between RagA and Nprl2 in *Drosophila* gut development. *Sheng Wu Gong Cheng Xue Bao* 2023;**39**:1747–58.
38. Min W, Fang D, Chen J *et al.* SpaMask: dual masking graph autoencoder with contrastive learning for spatial transcriptomics. *PLoS Comput Biol* 2025;**21**:e1012881. https://doi.org/10.1371/journal.pcbi.1012881
39. Huang L, Liu X, Zhu F *et al.* Improving cell-type composition inference in spatial transcriptomics with SpaDAMA. *PLoS Comput Biol* 2025;**21**:e1013354. https://doi.org/10.1371/journal.pcbi.1013354
40. Li X, Zhu F, Min W. SpaDiT: diffusion transformer for spatial gene expression prediction using scRNA-seq. *Brief Bioinform* 2024;**25**:bbae571. https://doi.org/10.1093/bib/bbae571
41. Min W, Shi Z, Zhang J *et al.* Multimodal contrastive learning for spatial gene expression prediction using histology images. *Brief Bioinform* 2024;**25**:bbae551. https://doi.org/10.1093/bib/bbae551
42. Zhao J, Min W. SpaICL: image-guided curriculum strategy-based graph contrastive learning for spatial transcriptomics clustering. *Brief Bioinform* 2025;**26**:bbaf433. https://doi.org/10.1093/bib/bbaf433
43. Min W, Xue S, Zhu F *et al.* SpaViT: self-supervised prediction of high-resolution spatial transcriptomics with vision transformer. *Tsinghua SciTechnol* 2025. https://doi.org/10.26599/TST.2025.9010087

Received: August 12, 2025. **Revised:** September 7, 2025. **Accepted:** September 10, 2025