Spatial genomics

Bihan Liu[a,b,c,d,e,*], **Yuxi Zhang**[f,*], **Chunlong Zhang**[f], **and Dan Tian**[a,b,c,d,e], [a]General Surgery Department, Beijing Friendship Hospital, Capital Medical University, Beijing, China; [b]National Clinical Research Center for Digestive Diseases, Beijing, China; [c]State Key Lab of Digestive Health, Beijing Friendship Hospital, Capital Medical University, Beijing, China; [d]Immunology Research Center for Oral and Systemic Health, Beijing Friendship Hospital, Capital Medical University, Beijing, China; [e]Beijing Laboratory of Oral Health, Capital Medical University School of Stomatology, Beijing, China; [f]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China

© 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Abstract

Advances in imaging, sequencing, and probe labeling have ushered in a new layer of information with spatial genomics. This approach describes the distribution and mutation of DNA and RNA from subcellular to tissue level. This chapter explores the advantages and challenges associated with various experimental approaches in spatial genomics. By integrating imaging and genomics data, bioinformatic algorithms can potentially reveal spatial distribution patterns and interactions within tissues. These findings can then be validated through robust experimental methods, ultimately leading to a deeper understanding of the fundamental mechanisms underlying disease progression and development.

Keywords

Spatial genomics; Spatial transcriptomics; Next generation sequencing; Bioinformatics; Nanostring digital spatial profiling; Visium spatial gene expression platform; BGI Stereo-seq; Spatial segmentation; Cell interaction; Cell communication

Key points

- Understanding spatial transcriptomics: rationale, experimental approaches, and bioinformatic analysis
- Capturing the spatial landscape: experimental approaches to collect spatial data
- Delving into the data: bioinformatic algorithms and tools for analysis
- Validating findings: experimental methods for functional confirmation

*These authors contributed equally in this chapter.

1 Introduction

1.1 Scope of spatial genomics

Spatial genomics encompasses a diverse array of techniques and methodologies aimed at elucidating the spatial organization and interactions of genetic material within the context of intact tissues. Combined with the single-cell sequencing, multiplex immunofluorescence, and bioinformatic algorithm, this burgeoning field merges concepts from genomics, transcriptomics, and imaging, offering unprecedented insights into the spatial heterogeneity of gene expression, genomic architecture, and cellular interactions within complex biological systems (Bressan et al., 2023). By preserving spatial information, spatial genomics approaches bridge the gap between traditional genomic analyses and histological examination, allowing researchers to investigate the spatial distribution of genomic signatures within tissue microenvironments.

The scope of spatial genomics extends across various biological scales and applications. Spatial genomics could be divided into two areas, RNA-related genomics and DNA-related genomics, while most research focused on RNA-related genomics, particularly spatial transcriptome (ST).

(1) RNA-related spatial genomics

RNA-related spatial genomics usually delves into the spatial production of mRNA, also known as spatial transcriptomics, within cells and tissues: mRNA localization entails the precise localization of RNA transcripts within distinct cells or subcellular compartments (Rao et al., 2021). Understanding the localization of mRNA molecules provides insights into their functional roles in essential cellular processes, including RNA processing, localization, and translation. Spatially regulated RNA localization plays a pivotal role in critical protein expression, cellular homeostasis, development, and response to environmental stimuli.

(2) DNA-related spatial genomics

DNA-related spatial genomics focuses on unraveling the chromatin architecture in the intranuclear level and nucleotide mutation distribution in the intercellular or tissue level. Chromatin architecture refers to the three-dimensional arrangement of chromatin within the nucleus and its impact on gene expression and genome function. Understanding chromatin architecture provides insights into gene regulation, genome stability, and cellular identity. Nucleotide mutation distribution indicates the development of cells, tissues and tumors and the function of the specific gene during these progresses (Lomakin et al., 2022). Spatial epigenomics approaches allow for the investigation of chromatin accessibility, DNA methylation patterns, and histone modifications. By preserving spatial information, these techniques uncover the DNA architecture of specific cell populations and spatially regulated gene regulatory networks, shedding light on the role of DNA status and mutation diversity in transcription, tissue homeostasis, and disease progression (Zhang et al., 2023).

In this chapter, we focused on the spatial transcriptomics, which is the typical and most concerned spatial genomics. The experiment and bioinformatic analysis strategy of spatial transcriptomics could be used as a reference in other types of spatial genomics.

1.2 Overview of spatial transcriptomics (ST)

To understand cellular systems and organ function in multicellular organisms, it's crucial to have comprehensive insights into their components and interactions. Under normal conditions, the overall systems maintain a dynamic equilibrium known as homeostasis. However, some immunological factors (e.g., pathogens) can disrupt this equilibrium and lead to different states. While single-cell RNA-sequencing (scRNA-seq) offers detailed gene expression profiles and heterogeneity within individual cells, ST provided spatial context further linking transcriptomes to specific cellular locations. This spatial information enables us to construct a map depicting cellular connections, factors influencing cell states, and their interactions with other systems at the level of organism. By analyzing these spatial relationships, we can uncover how different types of cells and genetic programs involved are interconnected and how they respond to changes within the surrounding environment.

ST technology originates from tissue in situ hybridization that detect single mRNA type using probes. Currently, the multiplexed in situ hybridization ST can detected thousands of mRNAs (over 18,000 protein-coding genes), many RNA types can be captured such as mRNAs, splice variants, lncRNAs, antisense RNAs and other RNAs. And multiple types of samples can be analyzed such as freshly frozen and formalin-fixed paraffin-embedded (FFPE) tissues. And the integration of ST data with other spatial and non-spatial techniques is playing a crucial role in deepening our understanding of the structural and functional states at the systemically level. Based on capture process and spatial store information, the ST methods can be classified into three types: in situ methods, spatial DNA-barcoding, and regional selection spatial transcriptomics methods. The in situ methods contained both spatial information and transcript identities in the acquired sample images, providing a location for each transcript detected at the level of single molecular. The second type of methods, spatial barcoding or DNA-barcoding is a collection of methods that capture mRNA from the tissue sample based on DNA-oligos with known barcoded capture regions, which incorporate the positional information with transcript sequences. The third type, regional selection barcoding is performed using photoactivatable markers and regional laser photoactivation. These methods included the TIVA tag and Zip-Seq tagged with barcode-oligo, and GeoMx system that detects the target RNAs using RNA hybridization probes. Similar as GeoMx, the laser capture microdissection captured transcripts from each cell or cell group, which is barcoded before sequencing or processed as sequencing libraries.

The evolving ST platforms have already demonstrated their significance in uncovering intricate details about cell heterogeneity and transcriptome interconnections within multicellular systems. In the meanwhile, the diverse platforms capable of generating vast

and complex omics datasets underscores the growing demand for advanced computational tools and expertise to effectively analyze this data. Thoughtful carefully data analysis are essential to enhance the efficiency and quality of findings in spatial transcriptomic research. The computational ST analysis and frameworks with optional standalone options are summarized and detailed shown in below.

1.3 Role of spatial genomics in tumor and immunologic diseases

1.3.1 Tumor microenvironment characterization

Spatial genomics techniques provide valuable insights into the tumor microenvironment, including the spatial distribution of immune cell populations, stromal components, and tumor cells (Hirz et al., 2023). By profiling spatially resolved molecular signatures within the tumor and its surrounding microenvironment, researchers can decipher the spatial heterogeneity of immune infiltration, immune cell interactions, and immune evasion mechanisms. This detailed characterization enhances our understanding of tumor-immune interactions and informs the development of immunotherapeutic strategies.

1.3.2 Immune checkpoint inhibitor response prediction

Spatial genomics enables the spatial mapping of immune checkpoint expression, such as PD-L1, CTLA-4, and others, within the tumor microenvironment. By correlating spatial expression patterns with treatment response, spatial genomics can predict the efficacy of immune checkpoint inhibitors (ICIs) and identify potential biomarkers for patient stratification. This personalized approach enhances the selection of patients likely to benefit from immunotherapy and improves treatment outcomes (Park et al., 2022).

1.3.3 Mechanisms of immune evasion

Through spatially resolved profiling of immune cell subsets and their spatial organization within tumors, spatial genomics elucidates the mechanisms of immune evasion employed by tumors. It reveals spatially restricted immune suppression, immune exclusion, and immunosuppressive niche formation within the tumor microenvironment. Understanding these spatial dynamics provides insights into therapeutic resistance mechanisms and guides the development of combination therapies targeting immune evasion pathways (Sun et al., 2021).

1.3.4 Spatial genomics in infectious and autoimmune diseases

Beyond cancer research, spatial genomics holds promise for studying infectious diseases, autoimmune disorders, and other pathological conditions. In infectious diseases, spatial genomics can delineate the spatial distribution of pathogens, host immune responses, and tissue damage within infected tissues (Byrne et al., 2016). This spatially resolved understanding informs the development of vaccines, antiviral therapies, and immunomodulatory interventions targeting specific disease compartments. In autoimmune disorders, spatial genomics enables the characterization of immune cell infiltration, tissue damage, and disease-specific molecular signatures within affected tissues. By uncovering spatially regulated immune pathways and dysfunctional tissue-immune interactions, spatial genomics facilitates the identification of novel therapeutic targets and the development of precision medicine approaches for autoimmune diseases (Zhao et al., 2015).

In summary, spatial genomics plays a pivotal role in advancing our understanding of tumor immunity and other diseases by providing spatially resolved insights into immune cell dynamics, tissue microenvironments, and disease pathogenesis. Leveraging spatially resolved molecular information holds immense potential for guiding therapeutic interventions, improving treatment outcomes, and advancing precision medicine initiatives across diverse disease contexts.

2 Body

2.1 Experimental approaches for ST data collection

2.1.1 Nanostring digital spatial profiling (DSP) technology

2.1.1.1 GeoMx DSP spatial genomics platform overview

The GeoMx Digital Spatial Profiler (DSP) is an advanced platform designed for spatial genomics analysis, providing researchers with valuable spatial and temporal information at the protein and RNA levels. This technology enables high-throughput profiling, capable of capturing digital readouts for proteins or RNA targets with spatial resolution in a single experiment. Through the utilization of Next-generation sequencing (NGS) technology or nCounter® barcoding technology, the GeoMx DSP provides an integrated approach to spatial transcriptomics, facilitating the comprehension of intricate biological processes and disease mechanisms (Baysoy et al., 2023).

Key components

(1) Digital Spatial Profiler (DSP): This instrumental component of the GeoMx DSP platform enables high-plex profiling of proteins and RNA within specific tissue regions, facilitating spatial transcriptomics.
(2) Programmable Digital Micromirror Device (DMD): The DMD is a crucial feature allowing flexibility and customization of regions of interest (ROI) during spatial profiling, enhancing the adaptability of the platform.

(3) Tissue Preparation with Immunofluorescence Biomarkers and DSP Probes: This step involves preparing tissue samples by labeling them with immunofluorescence biomarkers and DSP probes, ensuring targeted profiling.

Workflow

(1) Slide Preparation: Before selecting ROI regions, stain the slide by fluorescence-labeled antibodies and prepare the slide by applying a mixture of probes, including UV-photocleavable oligos, to the tissue section. These probes bind to specific RNA or protein targets within the tissue.
(2) ROI Selection: GeoMX DSP allows for various region-of-interest selection strategies, depending on the types of questions being investigated. Researchers can define ROI regions based on specific criteria such as cell types, morphological features, or other spatial patterns of interest.
(3) UV Exposure: After selecting the ROI regions, direct UV light onto these regions. This exposure cleaves the DSP barcodes from their probes or antibodies (Vandereyken et al., 2023).
(4) Count: Quantify RNA and protein expression levels with direct, digital counting of barcodes using the nCounter® Analysis System or sequencing on an Illumina platform.
(5) Data Analysis: Analyze the data with the provided DSP Data Analysis Suite (DSPDA), open-source tools, or Spatial Data Analysis Service (SDAS) to interpret spatial gene expression patterns.

Applications

(1) Tumor Microenvironment Analysis: Investigate the spatial heterogeneity of tumors, tumor-infiltrating immune cells, and stromal components within the tumor microenvironment.
(2) Neuroscience: Explore the spatial distribution of cell types, neuronal networks, and biomarkers in brain tissue, aiding in understanding neurodegenerative diseases and brain disorders.
(3) Immunology: Characterize immune cell infiltration patterns and spatial interactions in various disease states, providing valuable insights into autoimmune diseases, infectious diseases, and cancer immunotherapy.
(4) Biomarker Discovery: Identify spatially regulated biomarkers associated with disease progression, treatment response, and prognosis, facilitating the development of personalized medicine approaches.
(5) Drug Development: Assess drug distribution, target engagement, and pharmacodynamics within tissue samples, optimizing drug development pipelines and therapeutic strategies.

Advantages

(1) High-Plex Profiling: The GeoMX DSP allows for high-plex profiling at both the protein and RNA level, providing comprehensive spatial information.
(2) Spatially Resolved Analysis: DSP provides spatially resolved readouts, offering insights into the spatial distribution of biomolecules within tissues. This spatial information is crucial for understanding complex biological processes and disease mechanisms.
(3) Suitability for Various Sample Types: DSP is compatible with various sample types, including fresh frozen and formalin-fixed paraffin-embedded (FFPE) tissues, allowing for versatile applications across different experimental settings.
(4) Digital Readout: The technology provides a digital readout of targets, offering quantitative and precise measurement of biomolecule abundance within tissue samples.
(5) Barcoding Technology: Nanostring DSP utilizes nCounter® barcoding technology, enhancing assay sensitivity and reducing background noise, leading to accurate and reliable results.

2.1.2 10× Genomics spatial analysis technology

2.1.2.1 Visium spatial gene expression platform overview

Visium Spatial Gene Expression platform is developed by 10× Genomics for spatially resolved transcriptomics analysis. It enables researchers to investigate gene expression patterns within intact tissue sections while preserving spatial information. The Visium platform integrates innovative spatial barcoding technology with high-throughput RNA sequencing, allowing for comprehensive spatial profiling of gene expression across complex tissue microenvironments (Galeano Niño et al., 2022).

Key components

(1) Spatial Barcoding Technology: Visium utilizes spatially barcoded capture beads that are deposited onto tissue sections in a spatially defined pattern. Each bead carries a unique spatial barcode, allowing for precise mapping of gene expression data to specific spatial locations within the tissue.
(2) Tissue Section Preparation Kits: The Visium Spatial Gene Expression Solution includes tissue section preparation kits, which provide reagents and protocols for tissue fixation, permeabilization, and barcoding. These kits enable standardized and reproducible tissue processing for spatial transcriptomics analysis.
(3) Library Preparation Kits: Visium offers library preparation kits optimized for spatial RNA sequencing, including cDNA synthesis, library construction, and sequencing adapter ligation. These kits enable efficient capture and amplification of spatially barcoded RNA molecules for downstream sequencing.

(4) Spatial Transcriptomics Data Analysis Software: 10× Genomics provides dedicated software tools for the analysis of spatially resolved transcriptomics data generated using the Visium platform. These software packages facilitate data processing, spatial mapping, gene expression quantification, and visualization of spatial gene expression patterns.

Workflow

(1) Tissue Section Preparation: Tissue sections are prepared from fresh-frozen or fixed tissues using standard histological techniques. The sections are then mounted onto Visium spatial capture slides and subjected to tissue permeabilization and barcoding.
(2) Spatial Barcoding: Spatially barcoded capture beads are deposited onto the tissue sections, allowing for the capture and labeling of RNA molecules from spatially defined regions. Each bead captures RNA transcripts from a specific location within the tissue.
(3) Library Construction: Captured RNA molecules are reverse-transcribed into cDNA and amplified using PCR. Sequencing adapters are ligated to the amplified cDNA molecules, enabling them to be sequenced on next-generation sequencing platforms.
(4) RNA Sequencing and Data Analysis: The spatially barcoded cDNA libraries are sequenced using high-throughput RNA sequencing technologies. The resulting sequencing data are processed and analyzed using dedicated bioinformatics software to quantify gene expression levels and visualize spatial gene expression patterns within the tissue.

2.1.3 Space ranger software overview

Space Ranger is a bioinformatics software package developed by 10× Genomics specifically for the analysis of spatially resolved transcriptomics data generated using their Visium Spatial Gene Expression Solution. This software is designed to process raw sequencing data, perform quality control, align reads to a reference genome, and quantify gene expression at spatially defined locations within tissue sections.

Key features

(1) Data Processing: Space Ranger automates the processing of raw sequencing data, including demultiplexing, barcode processing, and read alignment. It efficiently handles large datasets generated from spatial transcriptomics experiments.
(2) Spatial Mapping: The software maps sequencing reads to spatially barcoded regions of interest on the tissue section, allowing for the generation of spatially resolved gene expression profiles. It assigns each read to its corresponding spatial location based on spatial barcodes.
(3) Gene Expression Quantification: Space Ranger quantifies gene expression levels at individual spatial locations, generating spatially resolved gene expression matrices. This enables the identification of spatially regulated genes and analysis of gene expression patterns within tissue microenvironments.
(4) Quality Control: The software performs quality control checks to assess data quality and sequencing depth across spatially resolved locations. It flags low-quality data points and provides metrics to evaluate the overall quality of the spatial transcriptomics dataset.
(5) Visualization and Analysis: Space Ranger offers visualization tools and interactive plots to visualize spatial gene expression patterns within tissue sections. It allows researchers to explore gene expression profiles, spatial clustering, and spatial relationships between genes of interest.
(6) Integration with Analysis Pipelines: Space Ranger seamlessly integrates with downstream analysis pipelines and bioinformatics tools for further exploration and interpretation of spatial transcriptomics data. It supports integration with popular analysis software such as Seurat, Scanpy, and Loupe Browser.

Advantages of 10× genomics spatial analysis technology

(1) High Resolution: Provides high spatial resolution, enabling the visualization of gene expression patterns at single-cell resolution within tissue sections.
(2) Multiplexed Analysis: Enables multiplexed analysis of gene expression in multiple regions of interest within the same tissue section.
(3) Comprehensive Data Analysis: Offers comprehensive bioinformatics tools and software platforms for data analysis, including visualization, clustering, and differential expression analysis.
(4) Compatibility with Various Sample Types: Compatible with diverse sample types, including fresh-frozen tissues, fixed tissues, and formalin-fixed paraffin-embedded (FFPE) tissues.
(5) Integration with Other Omics Technologies: Integrates seamlessly with other omics technologies, such as single-cell RNA sequencing (scRNA-seq) and chromatin accessibility profiling.

Application

(1) Tumor Heterogeneity: Facilitates the comprehensive analysis of intratumor heterogeneity, discerning various cell subpopulations, their interactions, and gene expression profiles throughout tumor domains.
(2) Tissue Morphology: Elucidates cellular spatial organization within tissues, unveiling critical spatial relationships crucial for physiological functions and pathological conditions.

(3) Host-Graft Response: Investigates host-graft interactions in transplantation scenarios, providing insights into cellular-level responses of host tissues to grafts or transplants, including the identification of immune cell infiltration and molecular alterations.
(4) Mechanisms of Tissue Development: Offers valuable insights into the dynamics of tissue development processes. Maps gene expression patterns during embryogenesis, organogenesis, and tissue remodeling, uncovering spatial cues governing cell fate determination.
(5) Response to Therapeutic Interventions: Evaluates the therapeutic efficacy at the tissue level, allowing the observation of drug impacts on specific cell populations within their native microenvironments. Additionally, it aids in identifying potential therapeutic targets.

2.1.4 BGI Stereo-seq technology

2.1.4.1 Spatial transcriptome-Stereo-seq technology overview

Stereo-seq, an innovative spatio-temporal omics technology developed independently by BGI Genomics, utilizes stereo chips to capture mRNA from tissue sections and reconstructs the spatial context using spatial barcodes (Coordinate ID, CID). This methodology lays a robust research foundation for elucidating the intricate interplay between gene expression, cell morphology, and the local environment. Stereo-seq represents a groundbreaking tool with Nanoscale Resolution, theoretically capable of achieving a 100% cell capture rate, leading to more informative and accurate cell clustering outcomes (Gong et al., 2024).

Key components

(1) Stereo Chips: These microfluidic devices are designed for capturing mRNA from tissue sections with high efficiency and spatial resolution. Stereo chips are crucial for preserving spatial information during the sequencing process.
(2) Spatial Barcodes (Coordinate ID, CID): Spatial barcodes are unique identifiers attached to each mRNA molecule, allowing researchers to reconstruct the spatial context of gene expression within tissues. These barcodes play a pivotal role in accurately mapping gene expression patterns.
(3) Fluorescent Imaging System: Utilized for visualizing the location of cell nuclei within tissue sections, enabling precise spatial mapping of gene expression at the single-cell level.

Workflow

(1) Tissue Preparation: Tissue samples are sectioned and mounted onto slides for subsequent mRNA capture.
(2) mRNA Capture: Stereo chips are used to capture mRNA molecules from tissue sections. Each mRNA molecule is tagged with a spatial barcode for later spatial reconstruction.
(3) Spatial Reconstruction: Spatial barcodes are decoded to reconstruct the spatial distribution of gene expression within the tissue sample.
(4) Sequencing: The mRNA molecules, along with their spatial barcodes, are sequenced to obtain spatially resolved transcriptomic data.
(5) Data Analysis: Advanced computational algorithms are applied to analyze the spatially resolved transcriptomic data, identifying spatially distinct gene expression patterns and cell types within the tissue.

2.1.4.2 DNBSEQ™ technology platforms overview

BGI Stereo-seq technology utilizes DNBSEQ™ sequencing platforms developed by MGI Tech for high-throughput genome sequencing. These platforms employ the innovative core technology known as DNBSEQ™, an abbreviation for DNA Nanoball Sequencing. DNBs, or DNA nanoballs, serve as crucial components generated during library preparation. These condensed structures are loaded onto flow cells, where the sequencing process occurs. In comparison to traditional sequencing platforms, DNBSEQ™ offers enhanced accuracy and efficiency in sequencing tasks.

Key features

(1) Increased Accuracy: DNBSEQ™ technology combines the low error accumulation of DNA Nanoballs (DNB) with advanced sequencing techniques, enhancing the accuracy of genomic sequencing.
(2) Ultra-High Throughput: DNBSEQ™ instruments offer ultra-high throughput capabilities, allowing for efficient and rapid sequencing of large volumes of genetic material.
(3) Wide Range of Applications: DNBSEQ™ platforms support various applications in genomics research, including whole-genome sequencing (WGS), whole-exome sequencing, and other genomic analyses.
(4) Innovative Core Technology: DNBSEQ™ includes various technologies related to DNA nanoballs, such as DNA single-strand circularization, DNB preparation technology, and Patterned Arrays, ensuring comprehensive and advanced sequencing capabilities

Applications

Stereo-seq technology has a wide range of applications in spatial genomics research:

(1) Cancer Research: Stereo-seq assists in identifying diverse cell types within tumors and analyzing their spatial distribution, offering valuable insights into tumor heterogeneity and interactions within the microenvironment. These are crucial for the progression of cancer research and the development of precision medicine.

(2) Developmental Biology: With its exceptional precision and resolution, Stereo-seq facilitates the examination of spatial gene expression patterns during development. This capability aids researchers in uncovering the intricate molecular mechanisms involved in embryogenesis and organogenesis.
(3) Neuroscience: Stereo-seq enables the generation of spatial gene expression maps in the brain, empowering researchers to explore neuronal diversity, circuitry, and alterations in gene expression associated with neurological disorders. Consequently, this technology promotes the comprehension of brain function and related disorders.
(4) Drug Discovery and Therapeutic Development: Stereo-seq technology facilitates spatially resolved drug screening assays, allowing researchers to assess drug penetration, target engagement, and therapeutic efficacy within specific tissue compartments.

Advantages

(1) Ultra-High Precision and Throughput: BGI Stereo-seq technology offers precision and throughput, enabling accurate mRNA capture from tissue sections with stereo chips.
(2) Unprecedented Spatial Resolution: Leveraging DNA Nanoball (DNB) technology, Stereo-seq offers researchers a powerful tool to explore spatial biology with an unprecedented field-of-view and resolution.
(3) Fine Resolution: Stereo-seq allows for fine resolution of 500 nm, enabling detailed examination of cellular structures and interactions.
(4) Wide Range of Applications: Facilitating the creation of multiple spatiotemporal cellular maps, Stereo-seq supports its use in diverse research areas such as cancer biology, developmental biology, and neuroscience.
(5) Transformative Influence: Stereo-seq, with its transformative influence, contributes to advancements in gene reading, writing, storage technologies, and disease prevention, positioning itself as a leading technology in spatial omics research.

2.2 Bioinformatic framework of spatial transcriptomics

2.2.1 In situ transcriptomic framework

The process of deriving cell-by-gene and cell-by-location matrices for ST analysis from a large set of raw in situ microscopy image data involves multiple computational steps. For instance, in a 69-bit 10,000 gene MERFISH experiment, each of the 256 tiled fields of views (FOV) requires 23 rounds of 3-color fluorescent images captured at 6 different focal z-planes, along with an additional single image at the fiducial bead z-plane. It results in 111,872 single-channel microscope images that need processing to extract 69-bit transcript codes and locations for constructing cellular transcriptome profiles. The more commonly used analysis steps with many ST types are shown in Section 2.2.4.

2.2.1.1 Preprocessing and spot registration

In raw images, the probed transcripts appear as signal spots. However, due to limitations in automated microscopy imaging and chromatic aberrations, these spots may not be consistently positioned across sequentially imaged FOVs. To address this, alignment of images within each FOV is performed using cross-correlation of fiducial marker peak signals or nuclear stains. Additionally, signal alignment and enhancement steps typically involve correction for chromatic aberration and illumination using control images specific to the microscope setup, as well as image deconvolution and background subtraction. To create a composite representation of the entire imaged sample, single or multi-channel FOVs can be stitched together using a process guided by the overlaps in the FOV tiles.

In the processed images, each positive spot or pixel serves as a potential transcript, possessing both a location within the pixel coordinate system and a unique transcript identifier based on the levels of sequential image channels. Spot identification involves detecting local maxima within the images and filtering out values based on a specified threshold. Following this, barcodes are extracted as location signal strings for subsequent decoding. The precise execution of these steps, including barcode decoding, error correction, and spot quality control, varies depending on the specific microscope setup and the chosen in situ ST methods. The decoded transcripts are then organized into a gene-by-location matrix, documenting the identifiers and 2D or 3D coordinates of each identified transcript within the designated coordinate system.

2.2.1.2 Spatial segmentation

To produce transcript profiles at the single-cell or other spatial unit level, the process involves segmentation and counting of detected transcripts to assign them to specific functional spatial units. Segmentation masks are created by outlining targeted features from images or algorithmically based on the spatial distribution of transcripts. Due to challenges in segmenting diverse biological samples, various computational methods have been developed, including manual segmentation based on marker threshold and automated machine learning strategies. In typical cell segmentation, distinct markers are used to divide the FOV into nuclear, cytoplasmic, and empty regions. Individual cells are delineated by selecting cell nuclei as centers and expanding the cytoplasmic area around them. Computational methods are then used to estimate the size and shape of cell boundaries, employing marker signals. Many machine learning-based segmentation tools are developed to predict spatial segmentation, including U-net, DeepCell, Mesmer, and CellPose. Recently, many other methods, such as SSAM, Baysor, and JSTA, are developed for some cases that only transcript signals are available. These algorithms utilized spatial transcript distributions or annotated transcript profiles to jointly

segment and annotate the cell type. Ultimately, the identified transcripts within spatial unit regions are quantified based on segmentation masks to create cell-by-gene and cell-by-location matrices for the downstream ST data analysis.

Analyzing ST data involves several crucial steps, including image processing, segmentation, decoding, and counting. These tasks can be performed using programming languages such as Python, R, or MATLAB, along with specialized image processing libraries or modules. Python-based tools such as Starfish and its offshoot SMART-Q provide scripts to create pipelines for processing raw microscope images obtained from various in situ ST techniques. These pipelines enable the generation of matrices representing the distribution of transcripts across cells and spatial locations. In addition to Starfish, there are other software options available for image processing, spot identification, and cell segmentation. PySpots, Cellpose, and FISH-quant are among the popular choices for these tasks in the Python ecosystem, while EBImage and imager are performed in R. For a comprehensive analysis of ST data, frameworks like Squidpy offer advanced features, including cell and nuclei segmentation. These tools enhance the efficiency and accuracy of analyzing complex ST datasets, enabling researchers to extract valuable insights from their experimental datasets.

2.2.2 Spatial barcoding ST methods

Spatial barcoding involves encoding transcript and location data in a DNA format, which is later converted into a digital format through sequencing for computational processing. This section displays the spatial-specific data analysis steps for spatial cell type pattern analysis based on raw sequencing data, and advanced analysis steps with common used ST dataset types are shown in Section 2.2.4.

2.2.2.1 Preprocessing and location matrix generation

The preprocessing steps for raw sequence data in spatial barcoding are similar as scRNA-seq analysis framework. In spatial transcriptomics, a gene to spot matrix is created instead of a gene to droplet barcode matrix in scRNA-seq data. Paired-end sequencing reads is used to build expression matrix, linking gene expression to spot position. The sequence alignment is performed based on annotated reference genome, and the exampled method is STAR aligner. Furthermore, a secondary alignment against a decoy genome can be performed to filter out undesired contaminating sequences. Each sequence pair includes a spatial barcode sequence and a UMI to remove PCR duplicates introduced during library preparation. Gene expression levels are then quantified from deduplicated, aligned reads, resulting in a spatial barcode by gene matrix. Commonly used tools for this purpose include STARsolo, bustools, ST Pipeline, Spaceranger count, and Slide-seq/drop-seq.

Each transcript and its corresponding location coordinate is connected by the spatial barcode-by-location data within the sample. These coordinates play a crucial role in several analytical tasks, including constructing spatial relationship graphs and grids for spot or cell interconnection analyses, assigning transcripts to cells following segmentation with subcellular resolution data, and facilitating visualizations and joint analyses of spots alongside various features on tissue images. The Visium platform, where barcode sequences and their positions in the spatial grid are predetermined, generates a barcode-by-location matrix for spatial analyses. For effective visualization and joint analysis of transcripts with tissue images, alignment between the barcode coordinate system and the tissue images is essential. Tools like Spaceranger count can detect spot positions from bright-field images, enabling spot-image alignment.

2.2.2.2 Estimation of the spot-wise cell type compositions

Spatial barcoding technology allows for the precise localization of transcripts at either multi- or subcellular spatial resolutions. In the context of multicellular resolution spatial barcoding, individual spots can contain transcripts originating from multiple cells. To unravel the cell compositions and regional enrichment of distinct cell types and stages within these spots, various computational methods were developed. SPOTlight employs non-negative matrix factorization (NMF) and SpatialDecon utilizes log-normal regression to deconvolute transcriptomic data. On the other hand, Cell2Location and Tangram leverages Bayesian model or deep learning framework for cell type resolution. Giotto, built on the ST-framework, adopts enrichment-based strategies for analyzing cell type compositions. Spot compositions are typically presented as proportions or probabilities of cell types present. Currently, the accuracy and resolution of cell type identification heavily rely on the compatibility of annotated reference transcriptome profiles derived from scRNA-seq or bulk RNA-seq datasets as the target sample reference. In the meanwhile, some reference-free deconvolution methods are also developed, such as conditional autoregressive-based deconvolution and latent Dirichlet allocation-based STdeconvolve. These methods are useful when reference scRNA-seq datasets are not available.

2.2.2.3 Spatial segmentation

Segmentation is the process of creating masks to assign detected transcripts to individual cells. In spatial barcoding data analysis, segmentation closely resembles the procedures used in in situ ST methods. This involves utilizing stained image data of the tissue sample along with gene signals. Typically, ST data includes image data with H&E staining, which helps identify nuclei locations and specific structural features. Additionally, fluorescent staining for different cellular markers can aid in guiding the segmentation process. However, segmenting tissues with densely packed cells currently remains a challenging task for ST analysis. Several factors should be considered when performin the segmentation process, such as sample characteristics and the platform information. Some studies have utilized basic grid segmentation techniques, however, high-resolution spatial barcoding transcriptomics data is not yet available currently.

2.2.3 Regional selection ST methods

Laser capture microdissection coupled with RNA-sequencing (LCM-seq) and digital spatial profiling (DSP) are techniques that utilize flexible regions of interest binning in ST. Each bin captures transcripts from a specific ROI and is barcoded to track its location within the sample. Once sequenced and processed, these bins are used similarly to the spots in spatial barcoding arrays, but with the flexibility for varying sizes, shapes, and discontinuous regions. Researchers have the freedom to determine the composition of each bin, including the option for a single or multiple cells' transcripts. The ROIs can be selected based on homogeneity in size and shape, cell type, cell marker.

During the ROI selection phase in GeoMx and LCM-seq analyses, cells or specific areas are often grouped together, a step dictated by either limitations in detection efficiency or the need for high throughput. This approach can lead to ROIs comprising cells at different developmental stages or of varied cell types without meticulous selection. Therefore, the analytical outcomes significantly hinge on the choices made during this selection stage. The analysis essentially involves comparing the transcriptomic profiles from ROIs of GeoMx and LCM-seq to identify genes that are expressed differently. For more tailored analyses, GeoMx data can be transitioned from its proprietary formats to a universally accessible spatial data format using the R-based GeoMx tools package. The filtration of ROIs and probes by leveraging quality control metrics related to sequencing, alignment, and the use of negative control probes can be performed. Also, the GeoMx tools includes the functional section such as normalization, dimensionality reduction, clustering, analyzing differential gene expression, and visualization display (UMAP, t-SNE, and volcano plots). For analyzing cell composition, GeoMx and LCM-seq's multicellular ROIs can be subjected to various deconvolution methods. Among these, the SpatialDecon R package stands out for its utilization of constrained log-normal regression for deconvolution, facilitating cell type prediction through established cell type signatures from RNAseq and scRNA-seq data. SpatialDecon can also craft novel cell profiles directly from isolated cell ROIs, employ tumor-specific cell type ROIs to filter out genes that could complicate the target cell profiles, thereby improving deconvolution accuracy for non-tumor cell types. To refine its estimates, SpatialDecon incorporates background noise from decoy probes and can integrate nuclei counts for a more accurate total cell count estimate. SpatialDecon also offers tools for neighborhood analysis, which involves modeling gene expression profiles and the regulation of gene activity in ROIs, using a method known as reverse deconvolution based on the estimated cell counts. While some ST-analysis tools can conduct based on computational studies in a compatible format, spatially advanced analyses leveraging the unique datasets of GeoMx and LCM-seq can be resolved for the ROIs.

2.2.4 Advanced solutions in ST data analysis

In ST data analysis, the overarching objectives revolve identifying, quantifying, and annotating spatial patterns across multiple biological scales, including multicellular, cellular, and molecular levels. Moreover, statistical analyses are crucial for assessing these features across different spatial dimensions. Many processing and analytical steps employed in the analysis of spatially resolved single-cell transcriptomes closely resemble those used in conventional scRNA-seq analysis. However, adjustments are necessary to address method-specific confounding factors and limitations inherent in ST analysis. In this section, we'll show the standard data analysis steps along with advanced computational methodologies tailored specifically for investigating biological phenomena in spatial data.

2.2.4.1 Available ST analysis frameworks

Popular scripting languages like R and Python, along with a range of open-source data analysis and visualization packages, are commonly utilized to construct ST data analysis and visualization pipelines. Recently, integrated analysis solutions have emerged, streamlining the process of conducting efficient and reproducible ST data analysis without requiring extensive programming or scripting expertise. And the common used frameworks such as Giotto, Seurat, STUtility, SPATA2, Squidpy, scvi-tools, stLearn, and GeoMx tools offer a comprehensive suite of functionalities for processing and analyzing ST data.

Giotto, SPATA2, Squidpy, scvi-tools, and stLearn frameworks have the capability to import diverse types of in situ and spatial barcoding ST datasets for analysis. STUtility primarily focuses on spatial barcoding data analysis, while Seurat extends its support to Vizgen MERFISH, GeoMx, and spatial barcoding data types. Additionally, GeoMx tools specialize in analyzing GeoMx DSP data. Notably, there are overlaps between some frameworks: Squidpy and scvi-tools incorporate features from the Scanpy package, while STUtility builds upon Seurat to enhance its ST data analysis functionalities. Giotto primarily operates within R but interfaces with Python modules through the reticulate interface. It offers a diverse set of tools for statistically detecting and analyzing various spatial patterns. Seurat also provides numerous functions for ST analysis and visualization, with seamless connectivity to other ST analysis packages in R. STUtility specializes in analyzing and visualizing sequential spatial sample layers in 3D, complementing general ST analysis tasks. Squidpy and scvi-tools, as modular Python frameworks, conduct statistical analysis and leverage Python's deep learning environments by providing standardized interfaces to higher-level machine learning packages. stLearn, another Python-based ST analysis framework, focuses on integrative deep learning-based analysis, spatial trajectory, and pseudotime analyses. SPATA2 is designed for trajectory and pseudotime analyses, integrating tools for enrichment analysis, and estimating genomic copy number variation (CNV).

2.2.4.2 Preprocessing and quality control

The ST data analysis frameworks rely on inputs including cell- or spot-by-gene matrix and cell- or spot-by-location matrix, optionally supplemented with image data of the sample tissue. All these frameworks mentioned above provide crucial

preprocessing capabilities for normalization, filtering, and dimensional reduction to ensure data quality and consistency. Preprocessing steps follow similar procedures to those used in scRNA-seq analysis, encompassing normalization across datasets, scaling, filtering out low-quality cells or spots, and excluding genes with low abundance based on total detected molecules. Specific scRNA-seq methods such as SCtransformation, SCnorm, or SCRAN can be used to remove unwanted variations. Notably, the stLearn package has introduced a novel deep learning-driven normalization approach tailored for spatial data, leveraging image data and neighboring spatial information to refine gene expression values. Typically, data undergoes dimensionality reduction through a variety of algorithms to facilitate complex tasks such as clustering and visualization of data variance into lower-dimensional spaces.

2.2.4.3 Finding cell type identity and patterns

An effective strategy for gaining insight into cellular patterns involves clustering and annotating cells or spots based on their transcriptomes, and then visualizing the spatial variation using tissue overlay plots. In addition to well-known clustering techniques for scRNA-seq data like K-means, hierarchical, and Louvain, specific algorithms tailored for ST data are available in frameworks like Seurat and STUtility. For example, Seurat employs modularity tuning in the KNN (k-nearest neighbor) distance graph for sample clustering, while STUtility utilizes NNMF dimensionality reduction. For ST data clustering, methods designed for scRNA-seq data clustering, including BayesSpace, SC3, IloReg, and SIMLR, can also be applied. And BayesSpace leverages spatial information for higher resolution delineation of spatial domains, and IloReg optimizes gene selection for clustering through a probabilistic feature extraction step before clustering.

Most ST data analysis frameworks offer one or more tools for differential gene expression analysis to profile marker genes and manually annotate cell types based on these markers. For example, scvi-tools provides two automated cell type annotation methods and the option to utilize scArches transfer learning models for dataset integration and annotation, even with different modality reference data. Many tools can be incorporated into ST analysis pipelines for cell type annotation, such as SingleR, CHETAH, and scGate in R, as well as CellO in Python. Moreover, there are methods available that leverage both tissue images and ST data for annotation purposes. SpaCell combines tissue images and ST data for cell type and disease stage annotation based on a deep learning solution. Identified cell types or unannotated cell clusters can be further analyzed visually and statistically for their spatial organization by plotting them on sample images and comparing them to existing tissue atlases provided by pathologists based on corresponding H&E stained tissue images. Incorporating regulon-based analysis can enhance cell stage clustering by uncovering coregulated genes and transcription factors that drive functional cell states. Presently, none of the existing ST data analysis frameworks include built-in gene regulatory network analysis capabilities. However, tools like SCENIC (Single-Cell rEgulatory Network Inference and Clustering) and pySCENIC are commonly utilized for detecting regulons in scRNA-seq data. And the users can SCENIC to cluster cells based on the cellular specific regulon activities.

2.2.4.4 Detecting spatial gene expression patterns

Genes exhibit spatial expression patterns at the molecular level, reflecting cell type-specific intrinsic programs and external influences on the local cell community or tissue microenvironment. By integrating cell molecular and organellar subcellular distribution data with ST information, a more comprehensive understanding of gene-phenotype associations within cells can be achieved. RNA velocity analysis, which involves examining exon–intron ratios in transcripts or inferring dynamics from nuclear to cytoplasmic gene expression ratios, can be employed to order cells along pseudotime trajectories for analyzing cell fate.

ST data analysis frameworks incorporate methods specifically designed to identify genes with spatial expression patterns using the spatial representations within the ST data. A common approach involves utilizing a spatial grid where the average gene expression within each grid box area is computed. Spatial barcode methods, which naturally present data in a grid-like structure due to their spatial organization, facilitate this analysis by arranging spots in a fixed-size grid with defined relative locations and neighboring spots. Each spot integrates gene expressions from its surrounding areas. In cases of high-resolution ST data, the dimensions of the grid boxes can be adjusted to capture expressions across varying scales, from subcellular to multicellular regions. Another prevalent strategy is to construct a spatial network where nodes represent cells and edges denote connections between cells. These edges may have associated weights based on factors such as physical distance or other connection metrics.

Various tools and techniques are developed to identify genes that exhibit spatial expression patterns, and Giotto's BinSpect method, SpatialDE, SPARK, and trendsceek are commonly used for this purpose. SpatialDE3 employs Gaussian process regression, SPARK utilizes generalized spatial linear models, and trendsceek relies on nonparametric marked point processes to detect spatial expression trends. Additionally, Seurat, SPATA2, Squidpy, and STUtility use different approaches such as autocorrelation-based methods and factor analysis for identifying genes with spatially correlated expression patterns.

Apart from identifying genes with specific spatial expression patterns, another key objective in ST data analysis is to detect spatial domains with coherent gene expression. Various computational techniques, including clustering, hidden Markov random field (HMRF) modeling, and cross-correlation analysis, are applied to define these spatial domains and summarize genes with similar expression patterns into metagenes. Giotto, SPARK, and SpatialDE utilize clustering methods, while MERINGUE uses cross-correlations for grouping expression patterns into spatial domains.

After statistically detecting spatial domains of coherent gene expression, these findings are often validated by comparing them with spatial domains identified from image data based on manual annotation. Recent methods like stLearn, SpaGCN, and SpaCell leverage joint analysis of molecular and image data to further refine spatial domain detection. stLearn employs a two-step clustering

method, SpaGCN integrates image RGB pixel data for tuning spatial expression graph weights, and SpaCell combines deep learning analysis of H&E stained images with spatial molecular data for cell type and disease stage classification based on ST datasets.

2.2.4.5 Cellular neighborhoods, and cell communications

Understanding how cells interact within tissues and organs is crucial for understanding biological systems and diseases. Recent ST advances, using data from single-cell RNA sequencing, have enabled researchers to study cell neighborhoods and molecular-level cell-cell interactions in detail. One key aspect of this analysis is identifying preferred cell-cell adjacencies and community compositions, which can indicate how different cell types interact. Tools like Giotto and Squidpy use statistical methods to compare observed cell-cell interactions with random baselines, helping to identify meaningful connections.

Another important aspect is predicting cell adhesion and signaling between cells, which is mediated by ligand-receptor pairs. Tools such as CellphoneDB, NicheNet, and CellChat provide curated lists of known ligand-receptor pairs for analysis. By calculating communication scores for these pairs, researchers can identify the most likely connections between specific cell types. Giotto implements a spatially-based method for calculating cell communications, and other methods such as stLearn and Squidpy display the enrichment analysis test for ligand-receptor pairs obtained from scRNA-seq datasets. And some novel methods are developed to identify ligand-receptor pairs at multiple length scales, such as Tensor-cell2cell and MISTy. Overall, these advancements in ST are helping researchers build comprehensive models of biological systems and identify potential targets for treating diseases.

2.2.4.6 Visualization and data interpretation

Most ST data analysis frameworks offer diverse options for graph production and spatial feature visualization, enhancing the exploration and interpretation of analyzed data. For instance, Giotto facilitates exploratory data browsing and enables the selection of specific data subsets for further analysis. SPATA2 offers an interactive tool for manual spatial trajectory drawing and annotation, enhancing user engagement in the analysis process. Squidpy employs its own image data container and integrates with Napari, a GPU-accelerated image analysis software, for advanced visualizations and image-based analysis. It allows the utilization of machine learning algorithms for feature extraction from various types of image data, such as H&E and fluorescent staining, aiding in cell and nuclei segmentation for subcellular spatial domain computations. Seurat and stLearn provide interactive viewing options and customizable visualization adjustments. STUtility offers tools for image trimming and both automatic and manual image alignment, facilitating the construction of rotating 3D stacks from spatial barcoding data of consecutive sample sections. While scvi-tools currently lacks integrated interactive spatial visualization, its use of anndata data container provides a workaround for importing analysis results into other software platforms for interactive visualization purposes. Overall, these diverse visualization capabilities empower researchers to explore and understand spatial transcriptomics data effectively.

2.3 Experimental verification methods

Spatial Genomics encompasses various experimental verification methods to validate the spatial organization and interactions of biomolecules within tissues. These methods are crucial for confirming the accuracy and reliability of spatially resolved data obtained from spatial transcriptomics or spatial proteomics techniques. Some commonly employed experimental verification methods include:

(1) Fluorescence in situ hybridization (FISH)
Fluorescence In Situ Hybridization (FISH) is a powerful molecular biology technique that allows for the visualization and localization of specific nucleic acid sequences within cells and tissues. It enables researchers to examine the spatial distribution of RNA or DNA molecules in their native cellular context, providing valuable insights into gene expression, chromosomal organization, and genomic alterations. These methods provide spatial context to gene expression data obtained from spatial transcriptomics techniques (Chrzanowska et al., 2020).

(2) Immunohistochemistry (IHC) and immunofluorescence (IF)
IHC and IF techniques involve the use of antibodies conjugated with fluorophores or enzymes to detect and visualize proteins in tissue sections. They can validate gene expression patterns at the protein level, corroborating the spatial distribution of transcripts identified through spatial transcriptomics analyses. They offer high specificity and sensitivity in detecting protein targets, allowing for precise localization within cellular compartments or tissue structures. These methods allow researchers to identify the spatial distribution of specific proteins and assess their expression levels within tissue microenvironments.

(3) Single-molecule RNA fluorescence in situ hybridization (smFISH)
smFISH is a high-resolution technique that enables the visualization and quantification of individual RNA molecules within cells. It offers single-molecule sensitivity and multiplexing capability, enabling the simultaneous detection of multiple transcripts within individual cells.

By employing multiple fluorescently labeled probes targeting different regions of the same RNA transcript, smFISH provides spatially resolved information about gene expression patterns at the single-cell level. It validates the expression of specific transcripts and elucidates their spatial distribution within tissue compartments, facilitating the characterization of cellular heterogeneity and spatially regulated gene expression (Diaz and Hermann, 2023).

(4) Spatial proteomics validation

Similar to spatial transcriptomics, the validation of spatial proteomics data involves comparing spatially resolved protein expression profiles with those obtained from traditional proteomic methods such as mass spectrometry or Western blotting.

Immunohistochemistry or immunofluorescence staining targeting specific proteins of interest serves as a valuable validation tool for confirming spatial protein distribution patterns (Guilliams et al., 2022).

(5) Functional assays

Functional assays, such as cell-cell interaction assays or organoid culture systems, can complement spatial genomics techniques by providing functional insights into the biological processes occurring within spatially defined tissue regions.

These assays help validate spatially resolved molecular interactions and provide mechanistic insights into the functional consequences of spatial organization within tissues.

Incorporating experimental verification methods into Spatial Genomics studies enhances the robustness and reliability of spatially resolved data, thereby facilitating a deeper understanding of tissue architecture, cellular interactions, and disease mechanisms.

3 Conclusion

Advancements in spatial transcriptomics are rapidly evolving, offering both methodological and computational solutions that cater to larger sample areas with higher resolutions. This progress results in the generation of spatial datasets of increasing complexity and size. The latest methodologies in spatial transcriptomics enable the collection of nearly complete transcriptomic datasets with subcellular resolution at the organ and systems level. These datasets contain a wealth of biological information regarding cell spatial relationships, which can enhance data interpretation and aid in identifying causal connections among genetic programs in multicellular systems. By harnessing these cutting-edge tools, researchers can delve deeper into understanding the intricate spatial organization of cellular processes and uncover new insights that advance our knowledge of complex biological systems.

The integration of statistical modeling, supervised and unsupervised machine learning techniques, and data fusion methods with existing knowledge of cellular pathways and regulatory networks holds significant promise for uncovering comprehensive sets of biological components involved in various processes. This approach also facilitates the dynamic modeling of cellular functions within complex multicellular systems.

However, the upcoming spatial datasets, expected to contain terabytes of raw data per sample and millions of individual transcripts with spatial information, pose computational challenges for traditional analysis methods. Yet, this influx of data presents a unique opportunity to innovate and develop new computational approaches tailored to the analysis of vast biological datasets.

Moreover, the importance of interactive data visualization and collaborative sharing is increasingly recognized in the analysis of complex biological data. While current spatial transcriptomics platforms may lack seamless support for standalone spatial data sharing and interactive visualization, there is a growing need to make data and analysis results accessible in user-friendly formats. By enabling interactive exploration and visualization of spatial datasets, researchers can enhance data understanding and promote broader engagement and collaboration within the scientific community and beyond.

By integrating spatial genomics data with complementary multi-omics datasets such as genomics, transcriptomics, proteomics, and metabolomics, researchers can achieve a thorough characterization of spatially resolved molecular profiles within tissues. This comprehensive approach enables a more holistic understanding of the spatial organization of molecular processes and interactions within biological systems (Vandereyken et al., 2023). Spatial genomics encompasses a wide range of techniques and applications focused on revealing the spatial organization and functional dynamics of the genome within intact tissues. By utilizing spatially resolved molecular data, spatial genomics has significant potential to enhance our comprehension of biological processes, disease mechanisms, and therapeutic approaches.

References

Baysoy A, Bai Z, Satija R, and Fan R (2023) The technological landscape and applications of single-cell multi-omics. *Nature Reviews. Molecular Cell Biology* 24: 695–713.

Bressan D, Battistoni G, and Hannon GJ (2023) The dawn of spatial omics. *Science* 381: Eabq4964.

Byrne AQ, Voyles J, Rios-Sotelo G, and Rosenblum EB (2016) Insights From Genomics Into Spatial And Temporal Variation In Batrachochytrium Dendrobatidis. *Progress in Molecular Biology and Translational Science* 142: 269–290.

Chrzanowska NM, Kowalewski J, and Lewandowska MA (2020) Use of fluorescence in situ hybridization (FISH) in diagnosis and tailored therapies in solid tumors. *Molecules* 25: 1864.

Diaz VD and Hermann BP (2023) Single-molecule fluorescence in situ hybridization for spatial detection of mRNAs in sections of mammalian testes. *Methods in Molecular Biology* 2656: 21–35.

Galeano Niño JL, Wu H, Lacourse KD, Kempchinsky AG, Baryiames A, Barber B, Futran N, Houlton J, Sather C, Sicinska E, Taylor A, Minot SS, Johnston CD, and Bullman S (2022) Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature* 611: 810–817.

Gong C, Li S, Wang L, Zhao F, Fang S, Yuan D, Zhao Z, He Q, Li M, Liu W, Li Z, Xie H, Liao S, Chen A, Zhang Y, Li Y, and Xu X (2024) Saw: An efficient and accurate data analysis workflow for Stereo-seq spatial transcriptomics. *Gigabyte*: 1–12.

Guilliams M, Bonnardel J, Haest B, Vanderborght B, Wagner C, Remmerie A, Bujko A, Martens L, Thoné T, Browaeys R, De Ponti FF, Vanneste B, Zwicker C, Svedberg FR, Vanhalewyn T, Gonçalves A, Lippens S, Devriendt B, Cox E, Ferrero G, Wittamer V, Willaert A, Kaptein SJF, Neyts J, Dallmeier K, Geldhof P, Casaert S, Deplancke B, Ten Dijke P, Hoorens A, Vanlander A, Berrevoet F, Van Nieuwenhove Y, Saeys Y, Saelens W, Van Vlierberghe H, Devisscher L, and Scott CL (2022) Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* 185(379–396): E38.

Hirz T, Mei S, Sarkar H, Kfoury Y, Wu S, Verhoeven BM, Subtelny AO, Zlatev DV, Wszolek MW, Salari K, Murray E, Chen F, Macosko EZ, Wu CL, Scadden DT, Dahl DM, Baryawno N, Saylor PJ, Kharchenko PV, and Sykes DB (2023) Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. *Nature Communications* 14: 663.

Lomakin A, Svedlund J, Strell C, Gataric M, Shmatko A, Rukhovich G, Park JS, Ju YS, Dentro S, Kleshchevnikov V, Vaskivskyi V, Li T, Bayraktar OA, Pinder S, Richardson AL, Santagata S, Campbell PJ, Russnes H, Gerstung M, Nilsson M, and Yates LR (2022) Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature* 611: 594–602.

Park S, Ock CY, Kim H, Pereira S, Park S, Ma M, Choi S, Kim S, Shin S, Aum BJ, Paeng K, Yoo D, Cha H, Park S, Suh KJ, Jung HA, Kim SH, Kim YJ, Sun JM, Chung JH, Ahn JS, Ahn MJ, Lee JS, Park K, Song SY, Bang YJ, Choi YL, Mok TS, and Lee SH (2022) Artificial intelligence-powered spatial analysis of tumor-infiltrating lymphocytes as complementary biomarker for immune checkpoint inhibition in non-small-cell lung cancer. *Journal of Clinical Oncology* 40: 1916–1928.

Rao A, Barkley D, França GS, and Yanai I (2021) Exploring tissue architecture using spatial transcriptomics. *Nature* 596: 211–220.

Sun YF, Wu L, Liu SP, Jiang MM, Hu B, Zhou KQ, Guo W, Xu Y, Zhong Y, Zhou XR, Zhang ZF, Liu G, Liu S, Shi YH, Ji Y, Du M, Li NN, Li GB, Zhao ZK, Huang XY, Xu LQ, Yu QC, Peng DH, Qiu SJ, Sun HC, Dean M, Wang XD, Chung WY, Dennison AR, Zhou J, Hou Y, Fan J, and Yang XR (2021) Dissecting spatial heterogeneity and the immune-evasion mechanism of CTCs by single-cell RNA-seq in hepatocellular carcinoma. *Nature Communications* 12: 4091.

Vandereyken K, Sifrim A, Thienpont B, and Voet T (2023) Methods and applications for single-cell and spatial multi-omics. *Nature Reviews. Genetics* 24: 494–515.

Zhang D, Deng Y, Kukanja P, Agirre E, Bartosovic M, Dong M, Ma C, Ma S, Su G, Bao S, Liu Y, Xiao Y, Rosoklija GB, Dwork AJ, Mann JJ, Leong KW, Boldrini M, Wang L, Haeussler M, Raphael BJ, Kluger Y, Castelo-Branco G, and Fan R (2023) Spatial epigenome-transcriptome co-profiling of mammalian tissues. *Nature* 616: 113–122.

Zhao M, Wang Z, Yung S, and Lu Q (2015) Epigenetic dynamics in immunity and autoimmunity. *The International Journal of Biochemistry & Cell Biology* 67: 65–74.