



HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2021 September 27.

Published in final edited form as:

Nature. 2021 August ; 596(7871): 211–220. doi:10.1038/s41586-021-03634-9.

Exploring tissue architecture using spatial transcriptomics

Anjali Rao^{1,*}, Dalia Barkley^{1,*}, Gustavo S. França¹, Itai Yanai^{1,2}

¹Institute for Computational Medicine, NYU Langone Health, New York, NY 10016, USA

²Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA.

Abstract

Deciphering the principles and mechanisms by which gene activity orchestrates complex cellular arrangements in multicellular organisms has far-reaching implications for research in the life sciences. Recent technological advancements in next-generation sequencing-based and imaging-based approaches have established the potential of spatial transcriptomics to measure expression levels of all or most genes systematically throughout tissue space, and have been adopted to generate biological insight in neuroscience, development, plant biology, and a range of diseases including cancer. Similar to datasets made possible by genomic sequencing and population health surveys, the large-scale atlases generated by this technology lend themselves to exploratory data analysis for hypothesis generation. Here, we review spatial transcriptomic technologies and describe the repertoire of operations available for paths of analysis of the resulting data. Spatial transcriptomics can also be deployed for hypothesis testing using experimental designs comparing timepoints or conditions - including genetic or environmental perturbations. Finally, spatial transcriptomic data is naturally amenable to integration with other data modalities providing an expandable framework for insight into tissue organization.

Many of the notable discoveries in the life sciences followed from the recognition that cellular organization within tissues is intimately linked to biological function. In developmental biology, central topics such as symmetry-breaking between daughter cells and cell fate decisions are based on spatial relationships between cells¹. In clinical settings, histopathology is often used as a conclusive diagnostic, precisely because diseases are characterized by abnormal spatial organization within tissues². Infectious and inflammatory processes can drastically change the cellular organization of tissues³. These discoveries were supported by methods in molecular biology - including *in situ* hybridization⁴ (ISH) and immunohistochemistry⁵ - that provided the ability to visualize biological processes more directly by mapping DNA, RNA and protein within tissues. However, these methods limit analysis to at most a handful of genes or proteins at a time.

itai.yanai@nyulangone.org.

*Equal first authors

Author contributions: AR, DB, and IY collectively wrote the review. All authors conceptualized the figures, which GSF then created. All authors edited and revised the manuscript.

Competing interests: The authors declare that they have no competing interests to the manuscript.

The ‘omics revolution has profoundly changed our ability to characterize cells. Instead of a few RNA or protein markers, new methods assay the full genome, transcriptome or proteome in cells⁶⁻⁹. This has led to the discovery of novel cell types and cell states and provided a more detailed understanding of biological processes in health and disease¹⁰⁻¹². Until recently however, these high-throughput techniques could not be applied *in situ*, resulting in the loss of information about spatial relationships among the catalogued populations of cells. To circumvent this limitation, early methods such as TomoSeq performed transcriptomics on serial slices to reconstruct a spatial axis¹³⁻¹⁶. Similarly, microdissection was used to manually isolate specific regions for scRNA-Seq, thus obtaining spatially-resolved information¹⁷⁻²³. Nanostring GeoMX digital spatial profiling was developed to capture targeted transcripts in manually selected regions of interest²⁴. To reconstruct spatial relationships between neighboring cells, creative methods rely on partial tissue dissociation²⁵, including ProximID²⁵⁻²⁷, PICseq²⁷ and ClumpSeq²⁸. In another approach, targeted mapping of a subset of genes can be used to ‘anchor’ single-cell RNA-Seq data²⁹⁻³³.

While these approaches enabled the reconstruction of tissue organization, they also highlighted the need for whole transcriptome, spatially-resolved methods. Over the past decade, technologies have emerged that bridge the gap between traditional approaches that retain spatial information (such as IF, ISH) and new methodologies with the ability to concurrently query the entire transcriptome in individual cells. The inception of this new approach of ‘spatial transcriptomics’ has facilitated novel discoveries in diverse fields from neuroscience to development to cancer. Here, we review common spatial transcriptomic technologies, discuss the principles of exploration of the data generated by these methods, examine the utility of spatial transcriptomics in different experimental designs, and highlight the promise of the technology for biological insights through integration with other modalities.

Spatial transcriptomics technologies

While key aspects of spatial transcriptomic technologies vary widely in terms of both the number of genes that can be probed as well as the size of tissue that can be assayed (Box 1), the methods reviewed here focus on technologies that allow transcriptome-level measurements across a tissue region. Spatial transcriptomics technologies are primarily categorized^{34,35} as (1) Next-Generation Sequencing (NGS)-based, encoding positional information onto transcripts before next-generation sequencing, and (2) imaging-based approaches, comprising *in situ* sequencing-based methods, where transcripts are amplified and sequenced in the tissue, and *in situ* hybridization-based methods, where imaging probes sequentially hybridized in the tissue³⁶⁻⁴⁰ (Figure 1a-c). This classification is not always clear-cut, and methods may incorporate elements from both categories. These diverse technologies can be seen as converging upon a gene expression matrix (Figure 1d) capturing the transcriptome at every spot (i.e. a pixel, a cell, or a group of cells).

Next-Generation Sequencing (NGS)-based approaches:

NGS-based approaches stem from the conceptual innovations of single-cell RNA-Seq methodologies and are contingent on the addition of a spatial barcode prior to library preparation³⁵ (Figure 1a). In 2016, Stahl *et al.* reported the first NGS-based method for spatial transcriptomics that enabled the capture of whole transcriptomes from tissue sections⁴¹. The central innovation was to capture poly-adenylated RNA on spatially-barcoded microarray slides prior to reverse transcription, ensuring that each transcript could be mapped back to its original spot using the unique positional molecular barcode. With each slide consisting of just over a thousand spots (100µm in diameter, 200µm center-center), large tissue areas could be investigated in an unbiased manner without selecting a region or importantly, a set of gene targets^{42,43}. The method was first demonstrated on the mouse olfactory bulb⁴¹, and has since been employed by several other groups⁴⁴⁻⁴⁷. 10x Genomics recently released an improved version of the technology called Visium, with increased resolution (55µm in diameter, 100µm center-center) and sensitivity (>10k transcripts per spot)⁴⁸. Many different fields have adopted this technology, including neuroscience⁴⁹, cancer biology^{47,50} and developmental biology⁵¹.

Slide-Seq, another NGS-based technology, uses randomly barcoded beads deposited onto a slide for mRNA capture⁵². Here, the position of each random barcode is obtained by *in situ* indexing. This method has achieved high-resolution (10µm) and recently increased sensitivity (500 transcripts per bead), which was found to be twice that of Visium for the same surface area⁵³. In parallel, high-definition spatial transcriptomics (HDST) also improved the resolution by replacing the glass slide with beads deposited in wells, similar to Slide-Seq⁵⁴. More recently, the DBiT-Seq⁵⁵ method has adopted microfluidics to apply polyT barcodes to the tissue section, while Stereo-seq uses randomly barcoded DNA nanoballs deposited in an array pattern to achieve nanoscale resolution^{56,57}. Seq-Scope has achieved subcellular resolution spatial barcoding and can be used to visualize nuclear and cytoplasmic transcripts⁵⁸. An innovative approach was adopted in Pixel-Seq where a polony-derived gel oligo array was used for RNA capture resulting in up to ~200 fold increase in resolution in comparison to existing methods⁵⁹.

Common to all NGS-based methods, the spatially-barcoded RNAs are collected and processed for sequencing. The barcode of each read is used to map the spatial position, while the rest of the sequencing read is mapped to the genome to identify the transcript of origin, collectively generating a gene expression matrix.

Imaging-based approaches:

Two main types of imaging-based approaches to spatial transcriptomics have been introduced: *in situ* sequencing- and *in situ* hybridization-based methods. *In situ* sequencing (ISS)-based methods directly read out the sequences of transcripts within the tissue. Specifically, the RNA is reverse transcribed, amplified by rolling circle amplification, and sequenced (Figure 1b). Ke *et al.*⁶⁰ first used this method by deploying targeted probes for the reverse transcription, followed by sequencing-by-ligation, and was implemented to study ~50 targeted genes in cancer^{60,61}, tuberculosis⁶², and brain development⁶³. Building upon this approach, STARMap incorporated advances in hydrogel chemistry, improved

padlock and primer design, and devised an error-robust sequencing-by-ligation method, and was thus able to profile thousands of genes in the mouse cortex⁶⁴. Other methods using sequencing-by-synthesis - as in BaristaSeq⁶⁵ and Barseq⁶⁶ - or sequencing-by-hybridization as in HybISS⁶⁷ - have led to increased read lengths, enabling higher throughput and cellular barcoding. Furthermore, *in situ* sequencing has been combined with cDNA extraction for NGS^{68,69}, highlighting the difficulty in classifying spatial transcriptomic methods as either NGS- or imaging-based. *In situ* sequencing also has the potential for untargeted profiling, as demonstrated by FISSEQ⁷⁰. Although the untargeted amplification can lead to optical crowding and lower sensitivity, the recently-developed ExSeq demonstrated that expansion microscopy can be used to perform untargeted *in situ* sequencing in tissues⁶⁹.

In situ hybridization (ISH) -based methods are the second category of imaging-based methods which build on *in situ* hybridization technologies, whereby a target sequence is detected by hybridization of a complementary fluorescent probe (Figure 1c). Initially limited in the number of distinguishable transcripts, innovations enabling the addition of sequential rounds of hybridization and imaging⁷¹ combined with barcoding have enabled substantial multiplexing. In MERFISH, successive rounds of hybridizations are imaged to detect the presence or absence of fluorescently labeled probes. The serial images are then decoded, using the error-robust barcode associated with each transcript identity⁷²⁻⁷⁴. MERFISH has been used at a wide range of scales, from transcript location within individual cells⁷⁵ to tissue-level spatial transcriptomics such as on the hypothalamic preoptic region⁷⁶. Another strategy to increase the number of distinguishable transcripts is the combination of colors into pseudocolors, as done in SeqFISH^{77,78}. Similar to MERFISH, this method can be applied to investigate intracellular organization⁷⁹ as well as to generate large maps, for example of the hippocampus⁷⁸. Both methods have improved considerably in the last few years, and are now able to detect ~10,000 genes at sub-cellular resolution^{75,80}. Ongoing efforts in the community aim to improve the sensitivity and scale of these methods^{34,81,82}.

For both ISS- and ISH-based methods, the image is processed to generate the gene expression matrix. To obtain a cell-level matrix, the image is segmented, either manually on small areas, or systematically using a computational approach. Watershed algorithms use DAPI-stained nuclei as seeds and identify cell borders as regions with low RNA density⁸³. Although these may not correspond to true physical boundaries, but rather to the limit between cells, they accomplish the task of assigning each mRNA to a cell. Alternatively, the data analysis can begin at the level of individual pixels, and incorporate the gene expression data to delineate cells⁸⁴⁻⁸⁶.

Spatial transcriptomics insights into development, physiology and disease

Since spatial transcriptomic technologies provide an unbiased picture of spatial composition, they have been used to generate tissue atlases, which provide a valuable resource as reference maps. The use of spatial transcriptomics to generate spatial atlases of the nervous system is of particular note: ST-based approaches have established detailed maps of the entire mouse brain⁴⁹, or of specific regions: visual cortex⁶⁴, primary motor cortex⁸⁷, middle temporal gyrus⁶⁷, hypothalamic pre-optic region⁷⁶, hippocampus^{69,78}, and cerebellum⁸⁸. Maynard *et al.* identified spatial patterns of known schizophrenia- and autism-related genes

in their analysis of the dorsolateral prefrontal cortex, that led to proposed mechanisms of genetic susceptibility to schizophrenia⁸⁹. Spatial transcriptomics was also used to identify genes and pathways in eight inflorescence tissue domains of *A. thaliana*⁹⁰.

In developmental biology, time-resolved spatial transcriptomics atlases have been useful to elucidate the spatial dynamics of heart development⁵¹, spermatogenesis⁹¹ and intestinal development⁹². Similarly, a comprehensive study of the human endometrium during the proliferative and secretory phases of the menstrual cycle identified a role for WNT and Notch signaling in regulating differentiation towards ciliated or secretory epithelial cells⁹³. In order to serve as effective resources for the research community, these atlases have been the focus of coordinated community efforts and are supported by the Human Cell Atlas project⁹⁴ and the Allen Institute for Brain Science⁹⁵.

Beyond normal development and physiology, spatial transcriptomics is well- positioned to study tissue disorganization in disease. Most prominently, spatial transcriptomics has enabled the identification of the mechanisms at play in cancer, where the tissue structure underlying normal physiological function is altered^{44,50,96-99}. With the increasing recognition of the importance of the tumor microenvironment, spatial transcriptomics has been used to address its relationship to cancer cells adopting different states^{45,46,69,98}. In particular, spatial transcriptomics enables the study of the molecular features across the cancer and normal tissue boundaries. For example, an immunomodulatory cancer cell state was revealed in skin squamous cell carcinoma⁴⁷. Spatial transcriptomics has also provided insights on the mechanisms of tissue dysregulation in neurodegenerative disorders - including Alzheimer's disease^{100,101} and amyotrophic lateral sclerosis¹⁰², infectious and inflammatory processes - such as leprosy¹⁰³, influenza¹⁰⁴ and sepsis¹⁰⁵, and rheumatological diseases - including rheumatoid and spondyloarthritis^{106,107}.

Spatial transcriptomics-enabled exploratory data analysis

The spatial transcriptomic technologies result in a gene expression matrix, which can be analyzed both to test existing hypotheses and to generate new observations through exploratory analysis. Given the complexity and high dimensionality of a spatial transcriptomic dataset, novel insights can arise from adopting a mindset open to finding unexpected relationships by data analysis. In this exploratory mode of data analysis - championed by John Tukey¹⁰⁸ - the result of one analysis guides the choice of the next, analogous to the way in which the result of a bench experiment guides the design of the next experiment. This is not to say that prior knowledge and hypotheses are ignored; rather that they are used to interpret results and direct the analyses. Thus, there is no predefined protocol in exploratory data analysis - no set pipeline for how to study a spatial transcriptomic dataset. Instead, there is a particular logic for how the data can be examined and a recognition of possible outcomes with each analysis^{109,110}.

Analyzing spatial transcriptomic data often requires the exclusion of low quality data and initial transformations on the gene expression matrix to increase the signal-to-noise ratio, which can be performed using analysis packages such as Giotto¹¹¹, Seurat^{112,113}, STUtility¹¹⁴, and STLearn¹¹⁵. The total number of transcripts detected in a spot provides

a first indication of the technical and biological attributes of the data. A relatively low number of transcripts per spot may indicate a technical artifact, such as insufficient permeabilization in certain regions, or a difference in cell density in the case of NGS-based methods. Alternatively, variation can arise from biological sources, such differences in transcriptional activity between cell types, or the presence of dying or necrotic cells, and this signal may confound downstream analyses. Smoothing algorithms can be applied to the data to increase sensitivity and to remove unwanted sources of technical and biological variation. Based on the premise that information can be shared between neighboring spots, averaging gene expression between physically adjacent spots in a moving window along the spatial coordinates can reduce noise⁴⁷. To compare the expression of a gene across spots, transcriptomes are often normalized by dividing by the total number of transcripts (TPM) or using regularized negative binomial regression¹¹⁶. Similarly, comparisons across genes are aided by scaling the data to have the same mean and variance across spots (z-score).

The normalized gene expression matrix provides the basis for initial observations at the level of individual genes or spots (Figure 1). Revealing structure in the data, such as cell type properties or coherent gene modules, requires further processing of the matrix. We distinguish five classes of operations that have been used to study spatial transcriptomic data, though more operations will undoubtedly be devised (Figure 2a). While applying any one operation to the data may not immediately lead to insight, using the operators serially based on the interpretation of the results at each stage can generate a ‘path’ to a result (Figure 2b).

Cluster.

The clustering operation reveals structure in the data, most basically defining sets of spots with similar transcriptomes or orthogonally, identifying genes with similar expression patterns across the spots. Similarity between spots can be calculated directly between transcriptomes using correlation or euclidean distance, or after dimensionality reduction such as PCA, tSNE and UMAP^{117,118}. These similarities are then used to cluster spots, for example using k-means, Louvain or hierarchical clustering¹¹⁹. These clusters may correspond to distinct regions or cell types in the tissue of study, which can then be annotated (see ‘Characterize’). In a study on gingivitis, spots clustered according to whether they were epithelial, connective or inflammatory¹²⁰. Clustering methods were also used to describe the tissue composition on sections of the plant *A. thaliana*, revealing four groups of spots corresponding to stem, meristematic area, flower reproductive organs, and sepals and petals⁹⁰.

Gene clustering, using the same approach, can identify co-expressed gene modules corresponding to a cell type or cell state¹¹¹. In spatial transcriptomic data from the cerebellum for example, clustering of genes identified two modules of spatially correlated genes in Purkinje cells⁵². Methods to cluster genes and spots simultaneously have also been used, including Non-negative Matrix Factorization (NMF)^{121,122} or factor analysis⁹⁶, where the gene expression matrix is factorized to reveal the underlying structure in spot clusters and gene modules. In prostate tumor samples, this revealed sets of spots and genes corresponding to cancer, stroma, and inflammation⁹⁶. Currently, clustering methods

focusing on the specific features of spatial transcriptomics are being developed, such as BayesSpace¹²³.

Select.

Typical spatial transcriptomic datasets contain more biological information than can be meaningfully interpreted by any single analysis. Therefore, it is usually appropriate to select a region of interest, for example a specific layer in the brain^{52,53}, or the interface between tumor and microenvironment^{87,124}. Orthogonally, one may focus the analysis on context-specific genes, either chosen *a priori* from biological knowledge - most notably in imaging-based methods which do not yet cover the whole transcriptome - or chosen from the dataset itself - by identifying highly variable genes for example. Gene selection methods abound, and those tailored to spatial transcriptomic data attempt to identify genes with high variance and whose expression is not random across the tissue. Genes can be scored according to their spatial autocorrelation (using Moran's I or Geary's C)¹²⁵, neighbor enrichment (for example, in BinSpect)¹¹¹ or entropy (for example, in Haystack)¹²⁶. Trendsceek¹²⁷ uses a marked point processes approach¹²⁸ and is able to identify hotspots, streaks, and gradients of expression. SpatialDE decomposes a given gene's expression variability into spatial and nonspatial components using Gaussian process regression¹²⁹, and a similar approach was extended upon in SPARK¹³⁰. Cancer-specific metabolic vulnerabilities were thus characterized by identifying spatially variable genes in prostate cancer⁹⁷.

Score.

While the genes and spots are the primary data observations of spatial transcriptomics, the underlying biology is such that genes are co-expressed as modules, and that spot transcriptomes reflect a finite set of cell types and states. This is the premise of the scoring function, which is used to summarize a cluster of similar spots as a single gene expression profile, or - orthogonally - a coherent set of genes as a single pattern. Summarizing in this way can identify functional properties - for example, a stress response state or infiltrating macrophages that are spatially organized within a tumor - which might not be detectable when analyzing spots or genes individually. Scoring can be done simply by averaging the values of the set, or by scoring the expression relative to a null model as implemented in the Seurat workflow¹¹³. In the brain for example, Moffitt *et al.* generated average cell type expression profiles to compare spatial transcriptomics and scRNA-Seq clusters⁷⁶. In melanoma, spots were scored according to their expression of previously established gene sets corresponding to cancer cell states⁴⁵ or to Gene Ontology terms¹²⁴.

Characterize.

The objects identified by operations on spatial transcriptomic data - clusters of spots and sets of genes - must be characterized for biological understanding and interpretation. For this, integration with other data sources and with other prior knowledge is essential. A cluster of spots may be characterized manually when it matches a histological region, as was done in MERFISH to annotate individual cell types in the brain⁷⁶ and in pancreatic cancer samples to annotate normal and malignant regions of the tumor⁴⁶. A cluster may also be annotated indirectly by identifying a set of marker genes and characterizing those. Specifically, a gene set can be characterized by quantifying its overlap with an annotated gene set. This is the

basis of the Multimodal Intersection Analysis (MIA) introduced by Moncada *et al.*⁴⁶, and of Gene Set Enrichment Analysis (GSEA) which queries for enrichment with functional groups obtained from Gene Ontology, KEGG, Hallmarks, and other databases¹³¹⁻¹³³.

Because NGS-based spatial transcriptomics is not at single-cell resolution, much attention has been given to the problem of inferring the cell type composition of each spot (deconvolution), which is an important step in building detailed organ atlases^{51,93}. Most methods achieve this by integrating single-cell data, either generated from the same sample (paired) or from a similar sample or database (unpaired). This integration helps to overcome the limitations of single-cell RNA-Seq – which lacks spatial information – and NGS-based spatial transcriptomics – which is not at single-cell resolution. The SPOTLight method uses non-negative linear regression on the spatial transcriptomic data using the NMF factors derived from single-cell to infer spot cell type composition¹³⁴. Similarly, NMF regression (NMFref) is used in SlideSeq⁵². Probability-based methods such as Stereoscope¹³⁵, Cell2location¹³⁶, and RSTG¹³⁷, as well as graph-based¹³⁸ and deep-learning based such as Tangram¹³⁹ have been introduced. In Stereoscope¹³⁵, the cell type parameters are assigned by maximum likelihood estimation on the single-cell and use those to estimate each spot composition. Cell2location¹³⁶ is similar to Stereoscope, but additionally attempts to infer the absolute number of cells per spot. DSTG¹³⁸ uses single-cell data to construct pseudo-spots, and then links real and pseudo spots in a graph of nearest neighbors. The spatialDWLS method borrows from methodologies previously used for bulk RNASeq deconvolution and applies cell type enrichment followed by a dampened weighted least squares method to determine spot composition¹⁴⁰.

ST methods with sub-cellular resolution face the inverse problem consisting of grouping spots into organelles or cells. Seq-scope makes use of transcript annotations as spliced, unspliced, or mitochondrial to define regions within cells⁵⁸. Recent approaches have been developed that use the local density of each RNA species to assign a cell type to each spot^{84,85}. Pci-Seq uses probabilistic cell typing and is able to identify cell types more efficiently in larger tissue areas^{23,86}. FICT, another method, integrates expression and neighbourhood information to assign cell types¹⁴¹. In the case of imaging-based methods, each DAPI-stained nucleus can be classified as a cell type according to its distance from marker gene RNAs⁸⁶.

Relate.

Given its systematic nature, spatial transcriptomics is well suited to identifying similarities, differences and relationships between populations of genes and tissue regions. Clusters of spots can be related by querying for expressed genes, spatial overlap, developmental or functional relationships. For example, Stickels *et al.*⁵³ identified genes that are differentially expressed between the proximal neuropil and the soma within the hippocampus using the different spots as replicates. Creative ways to relate the transcriptomes of clusters of spots are borrowed from those originally developed for scRNA-Seq. RNA Velocity¹⁴² makes use of the unspliced transcripts to infer how spots are related to each other in time, and was applied in the cortex to map the dynamics of neuro-development⁵³. RNA-Seq-based Copy-number variation inference identifies chromosomal aneuploidies, which can be used to

distinguish malignant from non-malignant spots, and also identify distinct subclones^{143,144}. When two sets of spots are spatially adjacent, potential modes of interaction¹⁴⁵ between the cells can be proposed by examining their paired receptors and ligands¹¹¹ using known databases such as CellPhoneDB^{47,93,146} or NicheNet¹⁴⁷.

ST for hypothesis generation and testing

ST atlases of healthy or diseased tissues naturally lend themselves to unbiased exploration and hypothesis generation^{51,93}. Even spatial transcriptomic datasets designed to study a specific biological process, such as time-course studies or perturbation experiments, can be explored to reveal unexpected changes and formulate new hypotheses¹⁰¹ (Figure 3). Thousands of spots or genes may be studied together, thereby exploiting the high dimensionality of the dataset to yield robust biological inferences. These observations - the presence of a cell type, a pattern of gene expression, or the co-localization of two cell states - may lead to a novel testable hypothesis. They should also be validated independently, for example by immunofluorescence⁴⁶ or *in situ* hybridization⁷⁶ (Figure 3a).

Alternatively, spatial transcriptomic data can be incorporated into a classical hypothesis-driven experimental design, whereby a sufficiently powered experiment is leveraged to test a well-defined prediction. Indeed, as spatial transcriptomic technology becomes more accessible, it is poised for use as a routine assay, on par with flow cytometry or RNA sequencing. Guided by experimental design, spatial transcriptomics can corroborate or falsify a hypothesis when used as a readout in a perturbation or time-course experiment. Each sample can be summarized by an individual datapoint, to be compared across replicates and conditions, necessitating that data be collected in sufficient numbers to ensure statistical rigor and power. Studies may incorporate spatial transcriptomics on several sections from the same sample to account for technical variability, or multiple biological replicates per condition. The hypothesis can further be tested in model systems, *in vitro* or *in vivo*, or in clinical data (Figure 3b).

Integration of spatial transcriptomics with other modalities

As the resolution and sensitivity of spatial transcriptomic technologies improve, integration with other data modalities can provide an opportunity for better tissue characterization using ST. While currently underutilized, the tissue image itself can be used to extract high resolution information, especially when combined with the vast knowledge acquired by the field of histopathology to manually identify and annotate regions². In particular, morphological features detected in the tissue such as cell shape or nucleus size can be directly incorporated in the analysis. In stLearn, spots with similar features are identified and spatial smoothing is improved by averaging across spots that are not only physically close but also similar in composition¹¹⁵. Another study improved the resolution of spatial transcriptomics gene expression data by fusing it with high-resolution histology image data¹⁴⁸. Deep learning has also been used to predict cell type annotations from gene expression and histology, outperforming annotations predicted from either modality alone¹⁴⁹. With the increase in transcriptomic data available for training, machine learning algorithms have also been used to predict gene expression from histopathology

images^{150,151}. Rather than relying on pre-defined morphological features, these algorithms improve their performance by decomposing the full image into “tiles”. Integration of spatial transcriptomics with such machine learning approaches may improve the interpretability of histopathology and its use in clinical decision making to guide treatment and inform prognosis.

At subcellular resolution, the spatial organization of chromatin may provide clues into the regulation of gene expression in various contexts. DNA seqFISH integrated with RNA seqFISH and multiplexed immunofluorescence revealed that active gene loci are located on the surface of nuclear bodies and zone interfaces in embryonic stem cells¹⁵². Integrating spatial transcriptomic datasets with high-throughput imaging of genome *in situ* and the spatial distribution of histone marks within a tissue will be extremely valuable¹⁵³⁻¹⁵⁵. Recently, spatial mapping of genome organization with concurrent DNA sequencing within intact tissue has been made feasible¹⁵⁶. This suggests that the goal of combining spatial genome sequencing with *in situ* transcriptomic profiling may be within reach, deepening our understanding how genome organization and function are encoded¹⁵⁵.

Augmenting gene expression data with a complementary modality like protein-co detection can also shed light into processes that spatial transcriptomics does not capture, such as post-translational modification and sub-cellular localization of proteins and their dysregulation in disease. Targeted protein co-detection performed alongside spatial transcriptomics can be achieved using immunostaining on the same tissue section, as enabled by Visium⁴⁸. A novel imaging cytometry based approach was used to simultaneously detect transcripts and proteins in breast cancer tissue samples¹⁵⁷. DBiT-Seq allows for the co-mapping of mRNA and proteins in the tissue using antibody-derived DNA tags, as is done in CITE-Seq¹⁵⁸. High-throughput spatial methods for protein detection such as MIBI, CODEX, t-cyCIF as well as mass spectrometry and barcode based approaches provide an unparalleled snapshot of the proteome within the tissue section¹⁵⁹⁻¹⁶⁴. Technological advances that allow the integration of these high-throughput proteomics methods with spatial transcriptomics will tremendously improve our ability to study tissue complexity.

Outlook

The spatial transcriptomics field is growing at an exponential pace, with daily releases of technologies and datasets. The challenges faced by current spatial transcriptomic methods - including the limits to resolution and sensitivity, as well as throughput and accessibility - are being rapidly overcome. Spatial transcriptomics methods are being made compatible with paraffin-embedded tissues, opening the door to retrospective analyses of samples collected over decades in biobanks^{48,70,165,166}. With future innovations, it may be possible to systematically assay larger tissue areas for the reconstruction of 3D organ- or organism-level atlases, and to visualize transcriptome-wide gene expression changes as they unfold over time. In addition to overcoming these technological challenges, future work will require the development of new computational tools and creative analytical thinking. Together, these will enable data exploration to identify ‘spatial patterns’ - a central feature of spatial transcriptomic datasets - and reveal insights into the underlying biology.

As we speculate about the future milestones of the field, the human genome project may serve as a useful parallel. The initial draft of the human genome was published in 2001^{167,168} and provided a reference to study the sources and consequences of genetic variation. However, the function and regulation of the different regions of the genome are still under active investigation. In spatial transcriptomics, future projects may similarly benefit from a reference from which to study distinct conditions. However, mapping the expression level of every gene in space will only be the first step to elucidating organizing principles of tissue biology. It is the coupling of these high-resolution cellular atlases with hypothesis-free inquiries that will enable new insight, and reveal the salient features of tissue architecture in physiology and disease.

A key challenge for the field will be to iteratively build a model of how multicellular spatial patterns emerge from cell-level properties. Independent of spatial transcriptomic technologies, implementing a simple principle - that each cell is overall most similar to its neighbors - was sufficient to recover complex spatial patterns in the *Drosophila* embryo¹⁶⁹. Building on this idea, the exploration of spatial transcriptomic datasets will enable us to uncover the fundamental principles that guide our modeling of tissue-level spatial organization and will facilitate the study of the mechanistic basis of these patterns and their consequences. These deeper biological insights will extend the level of understanding from simple tissues to more complex structures, including developing organisms and diseased tissues, bringing us closer to conquering the spatial frontier.

Acknowledgements:

We thank Felicia Kuperwaser, Andrew Pountain, Bo Xia, and other members of the Yanai lab, as well as Mark Phillips for critical reading and feedback. We thank the students of the exploratory data analysis course at NYU Langone Medical Center. IY was supported by grants from the NIH (R01AI143290) and the Lowenstein Foundation, and DB was supported by the NIH (F30CA257400). We also thank the three anonymous reviewers for important suggestions.

References:

1. Barresi MJF & Gilbert SF Developmental Biology. (Sinauer Associates, Incorporated, 2019).
2. Damjanov I & McCue PA Histopathology: A Color Atlas and Textbook. (Lippincott Williams & Wilkins, 1996).
3. Safai B & Good RA Immunodermatology. (Springer Science & Business Media, 2013).
4. Lehmann R & Tautz D In Situ Hybridization to RNA. in Methods in Cell Biology vol. 44 575–598 (Academic Press, 1994). [PubMed: 7535885]
5. Swanson PEFoundations of Immunohistochemistry. Am. J. Clin. Pathol90, 333–339 (1988). [PubMed: 3046324]
6. Mincarelli L, Lister A, Lipscombe J & Macaulay IC Defining Cell Identity with Single-Cell Omics. Proteomics 18, e1700312 (2018). [PubMed: 29644800]
7. Macaulay IC, Ponting CP & Voet T Single-Cell Multiomics: Multiple Measurements from Single Cells. Trends Genet. 33, 155–168 (2017). [PubMed: 28089370]
8. Liu XZ, Peng YZ & Shen ZA [Research advances on application of single-cell RNA sequencing in islet cell biology]. Zhonghua Shao Shang Za Zhi 36, 1208–1212 (2020). [PubMed: 33379859]
9. Xia B & Yanai I A periodic table of cell types. Development 146, (2019).
10. Zheng G, Xie Z-Y, Wang P, Wu Y-F & Shen H-Y Recent advances of single-cell RNA sequencing technology in mesenchymal stem cell research. World J. Stem Cells 12, 438–447 (2020). [PubMed: 32742561]

11. Lawson DA, Kessenbrock K, Davis RT, Pervolarakis N & Werb Z Tumour heterogeneity and metastasis at single-cell resolution. *Nat. Cell Biol* 20, 1349–1360 (2018). [PubMed: 30482943]
12. Papalexi E & Satija R Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol* 18, 35–45 (2018). [PubMed: 28787399]
13. Combs PA & Eisen MB Sequencing mRNA from cryo-sliced *Drosophila* embryos to determine genome-wide spatial patterns of gene expression. *PLoS One* 8, e71820 (2013). [PubMed: 23951250]
14. Junker JPet al.Genome-wide RNA Tomography in the zebrafish embryo. *Cell*159, 662–675 (2014). [PubMed: 25417113]
15. Lacraz GPAet al.Tomo-Seq Identifies SOX9 as a Key Regulator of Cardiac Fibrosis During Ischemic Injury. *Circulation*136, 1396–1409 (2017). [PubMed: 28724751]
16. van den Brink SCet al.Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature*582, 405–409 (2020). [PubMed: 32076263]
17. Nichterwitz Set al.Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat. Commun*7, 12139 (2016). [PubMed: 27387371]
18. Nichterwitz S, Benitez JA, Hoogstraaten R, Deng Q & Hedlund E LCM-Seq: A Method for Spatial Transcriptomic Profiling Using Laser Capture Microdissection Coupled with PolyA-Based RNA Sequencing. *Methods in Molecular Biology* 95–110 (2018) doi:10.1007/978-1-4939-7213-5_6.
19. Aguila Jet al.Spatial transcriptomics and in silico random pooling identify novel markers of vulnerable and resistant midbrain dopamine neurons. doi:10.1101/334417.
20. Moor AEet al.Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell*175, 1156–1167.e15 (2018). [PubMed: 30270040]
21. Chen Jet al.Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat. Protoc*12, 566–580 (2017). [PubMed: 28207000]
22. Tirosh Iet al.Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*352, 189–196 (2016). [PubMed: 27124452]
23. Peng Get al.Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature*572, 528–532 (2019). [PubMed: 31391582]
24. Geiss GKet al.Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol*26, 317–325 (2008). [PubMed: 18278033]
25. Halpern KBet al.Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. *Nat. Biotechnol*36, 962–970 (2018). [PubMed: 30222169]
26. Boisset J-Cet al.Mapping the physical network of cellular interactions. *Nat. Methods*15, 547–553 (2018). [PubMed: 29786092]
27. Giladi Aet al.Dissecting cellular crosstalk by sequencing physically interacting cells. *Nat. Biotechnol*38, 629–637 (2020). [PubMed: 32152598]
28. Manco Ret al.Clump sequencing exposes the spatial expression programs of intestinal secretory cells. *Cold Spring Harbor Laboratory* 2020.08.05.237917 (2020) doi:10.1101/2020.08.05.237917.
29. Satija R, Farrell JA, Gennert D, Schier AF & Regev A Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* vol. 33 495–502 (2015).
30. Achim Ket al.High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol*33, 503–509 (2015). [PubMed: 25867922]
31. Pettit J-Bet al.Identifying cell types from spatially referenced single-cell expression datasets. *PLoS Comput. Biol*10, e1003824 (2014). [PubMed: 25254363]
32. Karaiskos Net al.The embryo at single-cell transcriptome resolution. *Science*358, 194–199 (2017). [PubMed: 28860209]
33. Cang Z & Nie Q Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun* 11, 2084 (2020). [PubMed: 32350282]
34. Zhuang XSpatially resolved single-cell genomics and transcriptomics by imaging. *Nat. Methods*18, 18–22 (2021). [PubMed: 33408406]
35. Larsson L, Frisén J & Lundeberg J Spatially resolved transcriptomics adds a new dimension to genomics. *Nat. Methods* 18, 15–18 (2021). [PubMed: 33408402]

36. Crosetto N, Bienko M & van Oudenaarden A Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet* 16, 57–66 (2015). [PubMed: 25446315]
37. Moor AE & Itzkovitz S Spatial transcriptomics: paving the way for tissue-level systems biology. *Curr. Opin. Biotechnol* 46, 126–133 (2017). [PubMed: 28346891]
38. Asp M, Bergenstråhle J & Lundeberg J Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *Bioessays* 42, e1900221 (2020). [PubMed: 32363691]
39. Waylen LN, Nim HT, Martelotto LG & Ramialison M From whole-mount to single-cell spatial assessment of gene expression in 3D. *Commun Biol* 3, 602 (2020). [PubMed: 33097816]
40. Teves JM & Won KJ Mapping Cellular Coordinates through Advances in Spatial Transcriptomics Technology. *Mol. Cells* 43, 591–599 (2020). [PubMed: 32507771]
41. Ståhl P Let al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82 (2016). [PubMed: 27365449] This paper was the first to perform array-based spatial transcriptomics, using positional barcodes at a resolution of 200µm, and demonstrated the approach on the mouse olfactory bulb.
42. Jemt A et al. An automated approach to prepare tissue-derived spatially barcoded RNA-sequencing libraries. *Sci. Rep* 6, 37137 (2016). [PubMed: 27849009]
43. Salmén F et al. Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat. Protoc* 13, 2501–2534 (2018). [PubMed: 30353172]
44. Thrane K, Eriksson H, Maaskola J, Hansson J & Lundeberg J Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. *Cancer Res.* 78, 5970–5979 (2018). [PubMed: 30154148]
45. Baron M et al. The Stress-Like Cancer Cell State Is a Consistent Component of Tumorigenesis. *Cell Syst* (2020) doi:10.1016/j.cels.2020.08.018.
46. Moncada R et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol* 38, 333–342 (2020). [PubMed: 31932730] This paper performed array-based ST and introduced multimodal intersection analysis to study pancreatic cancer and identify cell type co-localizations.
47. Ji A Let al. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* 182, 497–514.e22 (2020). [PubMed: 32579974] This paper used array-based ST to identify genes at the leading edge of the tumor in squamous cell carcinoma.
48. Spatial Transcriptomics - 10x Genomics. 10x Genomics <https://www.10xgenomics.com/spatial-transcriptomics/>.
49. Ortiz C et al. Molecular atlas of the adult mouse brain. *Sci Adv* 6, eabb3446 (2020). [PubMed: 32637622]
50. Grauel A Let al. TGFβ-blockade uncovers stromal plasticity in tumors by revealing the existence of a subset of interferon-licensed fibroblasts. *Nat. Commun* 11, 6315 (2020). [PubMed: 33298926]
51. Asp M et al. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* 179, 1647–1660.e19 (2019). [PubMed: 31835037] Using an integrated approach of array-based ST, probabilistic cell typing by in situ sequencing and scRNAseq, this study built a comprehensive atlas of the developing human heart.
52. Rodriques S G et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363, 1463–1467 (2019). [PubMed: 30923225] This paper describes Slide-Seq, an array-based method with 10µm resolution, performed in the cerebellum and hippocampus.
53. Stickels R R et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol* (2020) doi:10.1038/s41587-020-0739-1.
54. Vickovic S et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* 16, 987–990 (2019). [PubMed: 31501547]
55. Liu Y et al. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* 183, 1665–1681.e18 (2020). [PubMed: 33188776]
56. Friedrich S & Sonhammer E L L Fusion transcript detection using spatial transcriptomics. *BMC Med. Genomics* 13, 110 (2020). [PubMed: 32753032]
57. Chen A et al. Large field of view-spatially resolved transcriptomics at nanoscale resolution. *bioRxiv* (2021) doi:10.1101/2021.01.17.427004.

58. Cho C-Set et al. Seq-Scope: Submicrometer-resolution spatial transcriptomics for single cell and subcellular studies. *bioRxiv* (2021) doi:10.1101/2021.01.25.427807.
59. Fu X et al. Continuous Polony Gels for Tissue Mapping with High Resolution and RNA Capture Efficiency. *Cold Spring Harbor Laboratory* 2021.03.17.435795 (2021) doi:10.1101/2021.03.17.435795.
60. Ke R et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* 10, 857–860 (2013). [PubMed: 23852452] This paper first performed in situ sequencing, mapping the expression of 31 transcripts using 4-base reads in breast cancer.
61. Darmanis S et al. Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep.* 21, 1399–1410 (2017). [PubMed: 29091775]
62. Carow B et al. Spatial and temporal localization of immune transcripts defines hallmarks and diversity in the tuberculosis granuloma. *Nat. Commun* 10, 1823 (2019). [PubMed: 31015452]
63. Tiklová K et al. Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nat. Commun* 10, 581 (2019). [PubMed: 30718509]
64. Wang X et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361, (2018).
65. Chen X, Sun Y-C, Church GM, Lee JH & Zador AM Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res.* 46, e22 (2018). [PubMed: 29190363]
66. Chen X et al. High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. *Cell* 179, 772–786.e19 (2019). [PubMed: 31626774]
67. Gyllborg D et al. Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res.* 48, e112 (2020). [PubMed: 32990747]
68. Fürth D, Hatini V & Lee JH In Situ Transcriptome Accessibility Sequencing (INSTA-seq). *Cold Spring Harbor Laboratory* 722819 (2019) doi:10.1101/722819.
69. Alon S et al. Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* doi:10.1126/science.aax2656.
70. Lee JH et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* 343, 1360–1363 (2014). [PubMed: 24578530] This paper introduced an untargeted in situ sequencing method, FISSEQ, that generated 30 base reads from 8102 genes demonstrated in human primary fibroblasts.
71. Codeluppi S et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* 15, 932–935 (2018). [PubMed: 30377364]
72. Chen KH, Boettiger AN, Moffitt JR, Wang S & Zhuang X RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348, aaa6090 (2015). [PubMed: 25858977]
73. Moffitt JR et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* 113, 11046–11051 (2016). [PubMed: 27625426]
74. Wang G, Moffitt JR & Zhuang X Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep* 8, 4847 (2018). [PubMed: 29555914]
75. Xia C, Babcock HP, Moffitt JR & Zhuang X Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Rep* 9, 7721 (2019). [PubMed: 31118500]
76. Moffitt JR et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 362, (2018). This paper combines MERFISH, an in situ hybridization-based method, and scRNAseq to investigate the hypothalamic pre-optic region.
77. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M & Cai L Single-cell in situ RNA profiling by sequential hybridization. *Nature methods* vol. 11 360–361 (2014). [PubMed: 24681720]
78. Shah S, Lubeck E, Zhou W & Cai L In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* 92, 342–357 (2016). [PubMed: 27764670]
79. Shah S et al. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* 174, 363–376.e16 (2018). [PubMed: 29887381] This paper establishes SeqFISH to study nascent RNAs using in situ hybridization in fibroblasts and embryonic stem cells.

80. Eng C-H Let al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* 568, 235–239 (2019). [PubMed: 30911168]
81. Crosetto N, Bienko M & van Oudenaarden A Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet* 16, 57–66 (2015). [PubMed: 25446315]
82. Asp M, Bergenstr hle J & Lundeberg J Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *Bioessays* 42, e1900221 (2020). [PubMed: 32363691]
83. Najman L & Schmitt M Watershed of a continuous function. *Signal Processing* (1994) doi:10.1016/0165-1684(94)90059-0.
84. Park J Segmentation-free Inference of Cell Types from in Situ Transcriptomics Data. (2020).
85. Littman Ret al. JSTA: joint cell segmentation and cell type annotation for spatial transcriptomics. doi:10.1101/2020.09.18.304147.
86. Qian X et al. Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* 17, 101–106 (2020). [PubMed: 31740815]
87. BRAIN Initiative Cell Census Network (BICCN) et al. A multimodal cell census and atlas of the mammalian primary motor cortex. *Cold Spring Harbor Laboratory* 2020.10.19.343129 (2020) doi:10.1101/2020.10.19.343129.
88. Kebschull J Met al. Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set. *Science* 370, (2020).
89. Maynard K Ret al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. doi:10.1101/2020.02.28.969931.
90. Giacomello S et al. Spatially resolved transcriptome profiling in model plant species. *Nat Plants* 3, 17061 (2017). [PubMed: 28481330]
91. Chen H et al. Dissecting Mammalian Spermatogenesis Using Spatial Transcriptomics. doi:10.1101/2020.10.17.343335.
92. Fawcner-Corbett D et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* 184, 810–826.e23 (2021). [PubMed: 33406409] Spatial transcriptomics was performed on the intestine at various stages of development, revealing how cell types are organized along the crypt-villus axis.
93. Garcia-Alonso L et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. doi:10.1101/2021.01.02.425073.
94. Rozenblatt-Rosen O et al. The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell* 181, 236–249 (2020). [PubMed: 32302568]
95. Lein E S et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176 (2007). [PubMed: 17151600]
96. Berglund E et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun* 9, 2419 (2018). [PubMed: 29925878]
97. Wang Y, Ma S & Ruzzo W L Spatial modeling of prostate cancer metabolic gene expression reveals extensive heterogeneity and selective vulnerabilities. *Sci. Rep* 10, 3490 (2020). [PubMed: 32103057]
98. Hwang W Let al. Single-nucleus and spatial transcriptomics of archival pancreatic cancer reveals multi-compartment reprogramming after neoadjuvant treatment. doi:10.1101/2020.08.25.267336.
99. Smith EA & Hodges HC The Spatial and Genomic Hierarchy of Tumor Ecosystems Revealed by Single-Cell Technologies. *Trends Cancer Res.* 5, 411–425 (2019).
100. Navarro J Fet al. Spatial Transcriptomics Reveals Genes Associated with Dysregulated Mitochondrial Functions and Stress Signaling in Alzheimer Disease. *iScience* 23, 101556 (2020). [PubMed: 33083725]
101. Chen W-T et al. Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer’s Disease. *Cell* 182, 976–991.e19 (2020). [PubMed: 32702314]
102. Maniatis S et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* 364, 89–93 (2019). [PubMed: 30948552]
103. Ma F et al. Single Cell and Spatial Transcriptomics Defines the Cellular Architecture of the Antimicrobial Response Network in Human Leprosy Granulomas. doi:10.1101/2020.12.01.406819.

104. Boyd DF et al. Exuberant fibroblast activity compromises lung function via ADAMTS4. *Nature* 587, 466–471 (2020). [PubMed: 33116313]
105. Janosevic Det et al. The orchestrated cellular and molecular responses of the kidney to endotoxin define a precise sepsis timeline. *Elife* 10, (2021).
106. Carlberg Ket et al. Exploring inflammatory signatures in arthritic joint biopsies with Spatial Transcriptomics. *Scientific Reports* vol. 9 (2019).
107. Vickovic Set et al. Three-dimensional spatial transcriptomics uncovers cell type dynamics in the rheumatoid arthritis synovium. doi:10.1101/2020.12.10.420463.
108. Tukey JW Exploratory Data Analysis. (1970). John Tukey established the field of exploratory data analysis as an approach to discover trends prior to testing for any particular model.
109. Yanai I & Lercher M What is the question? *Genome Biol.* 20, 1–5 (2019). [PubMed: 30606230]
110. Yanai I & Lercher M The data-hypothesis conversation. *Genome Biol.* 22, 1–6 (2021). [PubMed: 33397451]
111. Dries Ret et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* 22, 1–31 (2021). [PubMed: 33397451] Giotto enables analysis of spatial transcriptomics data, including normalization, identification of spatially variable genes and cell types, and ligand-receptor analysis.
112. Stuart Tet et al. Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888–1902.e21 (2019). [PubMed: 31178118]
113. Butler A, Hoffman P, Smibert P, Papalexi E & Satija R Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol* 36, 411–420 (2018). [PubMed: 29608179]
114. Bergenstråhle J, Larsson L & Lundeberg J Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* 21, 482 (2020). [PubMed: 32664861]
115. Pham Det et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. doi:10.1101/2020.05.31.125658.
116. Hafemeister C & Satija R Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 20, 296 (2019). [PubMed: 31870423]
117. Zhou B & Jin W Visualization of Single Cell RNA-Seq Data Using t-SNE in R. *Methods Mol. Biol* 2117, 159–167 (2020). [PubMed: 31960377]
118. Becht E et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol* (2018) doi:10.1038/nbt.4314.
119. Kiselev VY, Andrews TS & Hemberg M Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet* 20, 273–282 (2019). [PubMed: 30617341]
120. Lundmark A et al. Gene expression profiling of periodontitis-affected gingival tissue by spatial transcriptomics. *Sci. Rep* 8, 9370 (2018). [PubMed: 29921943]
121. Lee DD & Seung HS Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791 (1999). [PubMed: 10548103]
122. Gao Y & Church G Improving molecular cancer class discovery through sparse non-negative matrix factorization. *Bioinformatics* 21, 3970–3975 (2005). [PubMed: 16244221]
123. Zhao E et al. BayesSpace enables the robust characterization of spatial gene expression architecture in tissue sections at increased resolution. doi:10.1101/2020.09.04.283812.
124. Hunter MV, Moncada R, Weiss JM, Yanai I & White RM Spatial transcriptomics reveals the architecture of the tumor/microenvironment interface. doi:10.1101/2020.11.05.368753.
125. Moran PAP Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23 (1950). [PubMed: 15420245]
126. Vandenbon A & Diez D A clustering-independent method for finding differentially expressed genes in single-cell transcriptome data. *Nat. Commun* 11, 4318 (2020). [PubMed: 32859930]
127. Edsgård D, Johnsson P & Sandberg R Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods* 15, 339–342 (2018). [PubMed: 29553578] Trendsceek is a

- computational method for identifying spatially variable genes using marked point processes, and was tested on mouse olfactory bulb and hippocampus regions and human breast cancer tissue.
128. Illian J, Penttinen A, Stoyan H & Stoyan D Statistical Analysis and Modelling of Spatial Point Patterns. (John Wiley & Sons, 2008).
 129. Svensson V, Teichmann SA & Stegle O SpatialDE: identification of spatially variable genes. *Nat. Methods* 15, 343–346 (2018). [PubMed: 29553579] SpatialDE describes a statistical test for identifying genes with a spatial pattern using a non-parametric regression based approach.
 130. Sun S, Zhu J & Zhou X Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* 17, 193–200 (2020). [PubMed: 31988518] This paper reports a new computational method that uses a generalized linear spatial model to identify spatially variable genes from spatial transcriptomics datasets.
 131. Ashburner M et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet* 25, 25–29 (2000). [PubMed: 10802651]
 132. Subramanian A et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A* 102, 15545–15550 (2005). [PubMed: 16199517]
 133. Kanehisa M & Goto S KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30 (2000). [PubMed: 10592173]
 134. Elosua M, Nieto P, Mereu E, Gut I & Heyn H SPOTlight: Seeded NMF regression to Deconvolute Spatial Transcriptomics Spots with Single-Cell Transcriptomes. doi:10.1101/2020.06.03.131334.
 135. Andersson A et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun Biol* 3, 565 (2020). [PubMed: 33037292]
 136. Kleshchevnikov V et al. Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. doi:10.1101/2020.11.15.378125.
 137. Cable D et al. Robust decomposition of cell type mixtures in spatial transcriptomics. doi:10.1101/2020.05.07.082750.
 138. Su J & Song Q DSTG: Deconvoluting Spatial Transcriptomics Data through Graph-based Artificial Intelligence. doi:10.1101/2020.10.20.347195.
 139. Biancalani T et al. Deep learning and alignment of spatially-resolved whole transcriptomes of single cells in the mouse brain with Tangram. *bioRxiv* (2020) doi:10.1101/2020.08.29.272831.
 140. Dong R & Yuan G-C SpatialDWLS: accurate deconvolution of spatial transcriptomic data. doi:10.1101/2021.02.02.429429.
 141. Teng H, Yuan Y & Bar-Joseph Z Cell Type Assignments for Spatial Transcriptomics Data. *bioRxiv* 2021.02.25.432887 (2021) doi:10.1101/2021.02.25.432887.
 142. La Manno G et al. RNA velocity of single cells. *Nature* 560, 494–498 (2018). [PubMed: 30089906]
 143. Patel A P et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401 (2014). [PubMed: 24925914]
 144. Elyanow R, Zeira R, Land M & Raphael B STARCH: Copy number and clone inference from spatial transcriptomics data. *Phys. Biol* (2020) doi:10.1088/1478-3975/abbe99.
 145. Armingol E, Officer A, Harismendy O & Lewis NE Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet* 22, 71–88 (2021). [PubMed: 33168968]
 146. Efremova M, Vento-Tormo M, Teichmann SA & Vento-Tormo R CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc* 15, 1484–1506 (2020). [PubMed: 32103204]
 147. Browaeys R, Saelens W & Saeys Y NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* 17, 159–162 (2020). [PubMed: 31819264]
 148. Bergenstr hle L et al. Super-resolved spatial transcriptomics by deep data fusion. *Cold Spring Harbor Laboratory* 2020.02.28.963413 (2020) doi:10.1101/2020.02.28.963413.
 149. Tan X, Su A, Tran M & Nguyen Q SpaCell: integrating tissue morphology and spatial gene expression to predict disease cells. doi:10.1101/837211.
 150. He B et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* 4, 827–834 (2020). [PubMed: 32572199] This study develops a deep learning method to predict gene expression from histopathology images of breast cancer patients.

151. Levy-Jurgenson A, Tekpli X, Kristensen VN & Yakhini Z Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. *Sci. Rep* 10, 18802 (2020). [PubMed: 33139755]
152. Takei Yet al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 590, 344–350 (2021). [PubMed: 33505024]
153. Deng Yet al. Spatial Epigenome Sequencing at Tissue Scale and Cellular Level. doi:10.1101/2021.03.11.434985.
154. Nguyen HQ et al. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* 17, 822–832 (2020). [PubMed: 32719531]
155. Su J-H, Zheng P, Kinrot SS, Bintu B & Zhuang X Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* 182, 1641–1659.e26 (2020). [PubMed: 32822575]
156. Payne AC et al. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* 371, (2021).
157. Schulz Det al. Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. *Cell Syst* 6, 531 (2018). [PubMed: 29698648]
158. Stoeckius Met al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14, 865–868 (2017). [PubMed: 28759029]
159. Angelo Met al. Multiplexed ion beam imaging of human breast tumors. *Nat. Med* 20, 436–442 (2014). [PubMed: 24584119]
160. Keren Let al. A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* 174, 1373–1387.e19 (2018). [PubMed: 30193111]
161. Piehowski P Det al. Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100- μ m spatial resolution. *Nat. Commun* 11, 8 (2020). [PubMed: 31911630]
162. Lin J-R et al. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife* 7, (2018).
163. Goltsev Yet al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* 174, 968–981.e15 (2018). [PubMed: 30078711]
164. Kohman RE & Church GM Fluorescent in situ sequencing of DNA barcoded antibodies. *Cold Spring Harbor Laboratory* 2020.04.27.060624 (2020) doi:10.1101/2020.04.27.060624.
165. Liu Y, Enniful A, Deng Y & Fan R Spatial transcriptome sequencing of FFPE tissues at cellular level. *Cold Spring Harbor Laboratory* 2020.10.13.338475 (2020) doi:10.1101/2020.10.13.338475.
166. Nagarajan MB, Tentori AM, Zhang WC, Slack FJ & Doyle PS Spatially resolved and multiplexed MicroRNA quantification from tissue using nanoliter well arrays. *Microsyst Nanoeng* 6, 51 (2020). [PubMed: 32419951]
167. Lander E Set al. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001). [PubMed: 11237011]
168. Venter J Cet al. The sequence of the human genome. *Science* 291, 1304–1351 (2001). [PubMed: 11181995]
169. Nitzan M, Karaikos N, Friedman N & Rajewsky N Gene expression cartography. *Nature* 576, 132–137 (2019). [PubMed: 31748748] The NovaSparc method reconstructs spatial organization from single-cell RNA-Seq data by positing that cells are overall most similar to their neighbors in systems that are inherently symmetrical, and was demonstrated in liver lobules, intestinal villi, and whole *Drosophila* embryos.
170. Xia C, Fan J, Emanuel G, Hao J & Zhuang X Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. U. S. A* 116, 19490–19499 (2019). [PubMed: 31501331]
171. Joglekar A et al. A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat. Commun* 12, 463 (2021). [PubMed: 33469025]
172. Frieda K Let al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* 541, 107–111 (2017). [PubMed: 27869821]

173. Rodriques S, Get al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363, 1463–1467 (2019). [PubMed: 30923225]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Box 1:**Considerations for selecting a spatial transcriptomic method****Gene throughput:**

NGS-based methods are unbiased as they capture all polyadenylated transcripts, and are therefore well suited for exploring a new system. In contrast, ISH- and most ISS-based methods (with the exception of FISSEQ⁷⁰ and ExSeq^{69,70}) are targeted and require *a priori* knowledge of the genes of interest. Nonetheless, the throughput of these methods has increased in recent years, reaching 10,000 genes^{166,170}. Targeted spatial transcriptomic methods can also be used in conjunction with single-cell RNA-Seq, where genes of interest have already been identified and can then be located more precisely^{51,63}. In addition, probes for non-polyadenylated transcripts can be designed to query for other RNAs such as mature microRNAs and tRNAs¹⁷⁰.

Sequence information:

In NGS-based and ISS-based methods, the cDNA sequence itself is a read out, enabling the detection of splice isoforms^{58,56,171,56} as well as single nucleotide variants and point mutations⁶⁰. When integrated with the gene expression matrix, this data can assist with reconstructing a time-course - using RNA Velocity⁵³ or lineage tracing¹⁷².

Sensitivity:

ISH-based methods are highly sensitive, recently reaching 80% detection efficiency relative to the gold standard smFISH¹⁷⁰. Sensitivity of the NGS-based methods is significantly lower and remains inferior to scRNA-Seq, but is rapidly improving to ~100 unique transcripts per square μm^2 ^{54,58,59,173}. There is generally a tradeoff between sensitivity and gene throughput, as seen in the higher sensitivity of targeted ISS-based methods⁶⁴ relative to the unbiased methods⁷⁰.

Resolution:

The resolution of *in situ* methods is limited only by the optical diffraction limit, and with expansion microscopy has reached ~100nm^{80,170}. These methods are therefore well-suited to questions concerning sub-cellular organization. NGS-based methods are limited by the diameter of spots, but their resolution has rapidly increased since the original method⁴¹, recently reaching ~1 μm ^{58,59}.

Area size:

The *in situ* methods can span a wide range of sizes, although there is a tradeoff between tissue size and imaging time⁷³. The NGS-based methods on the other hand are standardized, with arrays on the order of ten square mm (6 for Visium), which may be inappropriate for smaller or larger samples.

Feasibility:

While these technologies are extremely powerful, there are obstacles to their widespread adoption, including access to single molecule imaging for *in situ* methods, as well as

manufacturing for the capture arrays. Commercialization has facilitated access in some cases.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

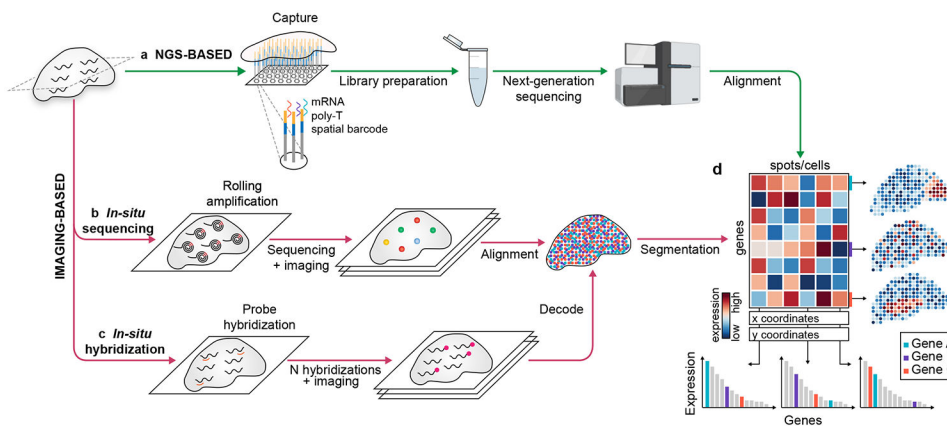


Fig. 1 |. The technologies of spatial transcriptomics provide a gene expression matrix. **a**, NGS-based spatial transcriptomic methods barcode transcripts according to their location in a lattice of spots. **b**, *In situ* sequencing approaches directly read out the transcript sequence within the tissue. **c**, *In situ* hybridization methods detect target sequences by hybridization of complementary fluorescent probes. **d**, The product of spatial transcriptomics is the gene expression matrix - where the rows and columns correspond to genes and locations.

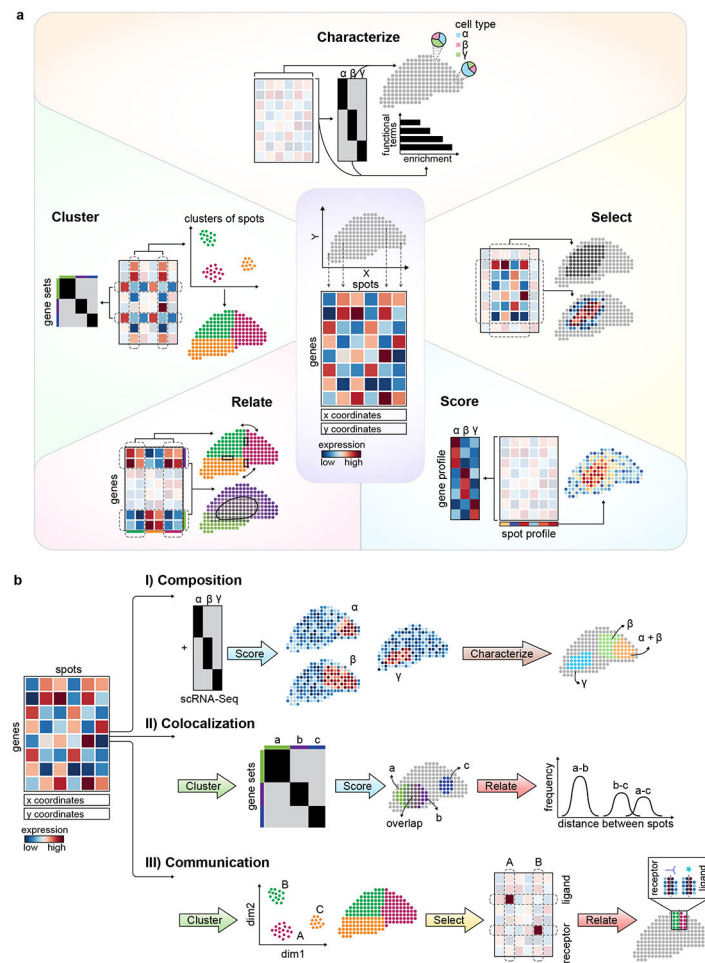


Fig. 2 | Exploratory data analysis using spatial transcriptomic datasets.

a. Schematic of exploratory data analysis operations of spatial transcriptomic datasets.

Characterize: Depicted are spots characterized to be composed of proportions of cell type α , β and γ and gene sets annotated with functional terms. **Cluster:** Clusters of spots are shown in a lower dimensional space and mapped onto the tissue, and co-expressed gene sets are shown within a gene-gene correlation matrix. **Select:** A subset of spots can be selected based on histological information, or a subset of spatially variable genes may be selected for analysis. **Relate:** The relationship between gene sets found to have a spatial overlap as well as adjacent spots and clusters can be examined using the relate operation. **Score:** Spots scored for gene set expression generate a spatial pattern, while gene profiles can be obtained by summarizing the expression of a subset of spots. **b.** Operational paths for analysis. **Composition:** Spots are scored for cell type-specific gene expression profiles from scRNA-Seq data and characterized to identify the composition of the tissue region. **Co-localization:** Co-varying genes are identified by clustering and spots are scored for the expression of these gene sets to identify a pattern of overlapping spatial expression. A co-localization is described by relating the distance between these spots. **Communication:** Transcriptionally similar spots are identified by clustering and characterized according to their resident cell types. A subset of receptor and ligand pairs are selected for analysis.

Characterize: Depicted are spots characterized to be composed of proportions of cell type α , β and γ and gene sets annotated with functional terms. **Cluster:** Clusters of spots are shown in a lower dimensional space and mapped onto the tissue, and co-expressed gene sets are shown within a gene-gene correlation matrix. **Select:** A subset of spots can be selected based on histological information, or a subset of spatially variable genes may be selected for analysis. **Relate:** The relationship between gene sets found to have a spatial overlap as well as adjacent spots and clusters can be examined using the relate operation. **Score:** Spots scored for gene set expression generate a spatial pattern, while gene profiles can be obtained by summarizing the expression of a subset of spots. **b.** Operational paths for analysis. **Composition:** Spots are scored for cell type-specific gene expression profiles from scRNA-Seq data and characterized to identify the composition of the tissue region. **Co-localization:** Co-varying genes are identified by clustering and spots are scored for the expression of these gene sets to identify a pattern of overlapping spatial expression. A co-localization is described by relating the distance between these spots. **Communication:** Transcriptionally similar spots are identified by clustering and characterized according to their resident cell types. A subset of receptor and ligand pairs are selected for analysis.

Receptors and ligands expressed in cell type α and cell type β , respectively, suggests a relationship between them.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

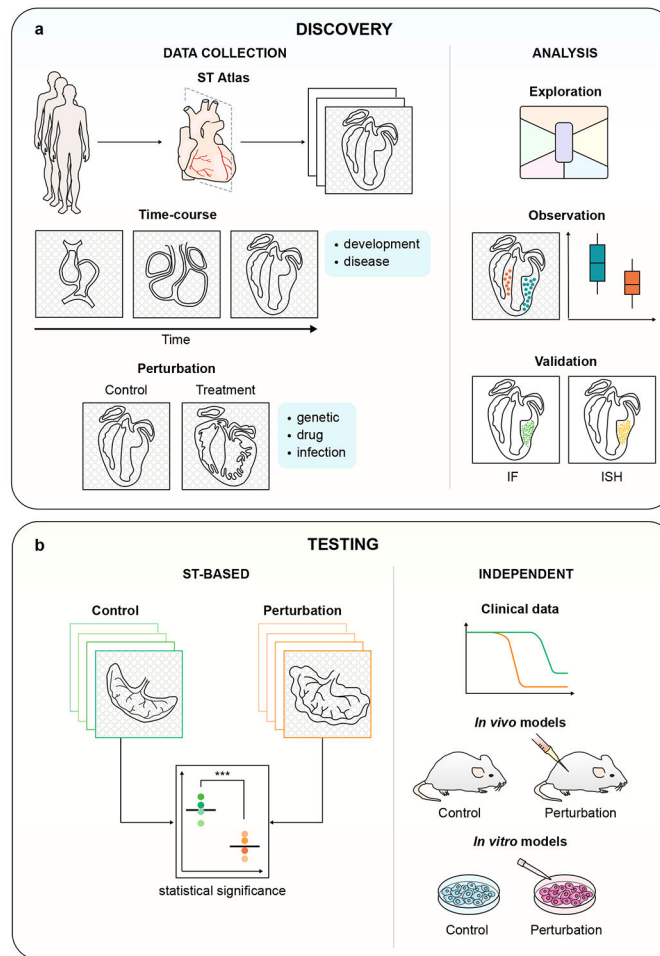


Fig. 3 | Hypothesis generation and testing using spatial transcriptomics.

a. Spatial transcriptomics can be used for hypothesis generation in various experimental contexts. Examples of spatial transcriptomic datasets include normal tissue (atlas), a developmental or disease time-course, and perturbation experiments (genetic, drug or infection). Following data collection, exploratory data analysis may generate observations - requiring validation - that lead to a hypothesis. **b.** Spatial transcriptomics for hypothesis testing. A well-powered experimental design that uses spatial transcriptomics can test formulated hypotheses. These can be further tested using clinical data, *in vivo* or *in vitro* models.