Exploring tissue architecture using spatial transcriptomics

https://doi.org/10.1038/s41586-021-03634-9

Received: 3 February 2021

Accepted: 11 May 2021

Published online: 11 August 2021

Check for updates

Anjali Rao[1,3], Dalia Barkley[1,3], Gustavo S. França[1] & Itai Yanai[1,2]✉

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 advances in next-generation sequencing- and imaging-based approaches have established the power of spatial transcriptomics to measure expression levels of all or most genes systematically throughout tissue space, and have been adopted to generate biological insights in neuroscience, development and plant biology as well as to investigate a range of disease contexts, 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 that compare time points or conditions—including genetic or environmental perturbations. Finally, spatial transcriptomic data are 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 in 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 cells[1]. In clinical settings, histopathology is often used as a conclusive diagnostic tool, precisely because many diseases are characterized by abnormal spatial organization within tissues[2]. Infectious and inflammatory processes can drastically change the cellular organization of tissues[3]. These discoveries were supported by methods in molecular biology—including in situ hybridization[4] (ISH) and immunohistochemistry[5]—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.

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 cells[6–9]. This has led to the discovery of novel cell types and cell states and has provided a more detailed understanding of biological processes in health and disease[10–12]. 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 performed transcriptomics on serial slices to reconstruct a spatial axis, such as serial microtomy sequencing (tomo-seq)[13–16]. Similarly, microdissection was used to manually isolate specific regions for single-cell RNA sequencing (scRNA-seq), thus providing spatially resolved information[17–23]. Nanostring GeoMX digital spatial profiling was developed to capture targeted transcripts in manually selected regions of interest[24]. To reconstruct spatial relationships between neighbouring cells, creative methods used partial tissue dissociation[25], including ProximID[26], cell sorting of physically interacting cells with single-cell RNA sequencing (PIC-seq)[27] and sequencing of cell clumps (ClumpSeq)[28]. In another approach, targeted mapping of a subset of genes was used to infer cell locations in whole-transcriptome scRNA-seq data[29–33].

While these approaches enabled the reconstruction of tissue organization, they also highlighted the need for spatially resolved methods that query the whole transcriptome. Over the past decade, technologies have emerged that bridge the gap between traditional approaches that retain spatial information (such as immunofluorescence or ISH) and new methodologies with the ability to concurrently query the entire transcriptome. The inception of this new approach of 'spatial transcriptomics' has facilitated novel discoveries in diverse fields, including neuroscience, development and 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 and the size

[1]Institute for Computational Medicine, NYU Langone Health, New York, NY, USA. [2]Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY, USA. [3]These authors contributed equally: Anjali Rao, Dalia Barkley. ✉e-mail: itai.yanai@nyulangone.org

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. By contrast, ISH- and most ISS-based methods (with the exception of FISSEQ[70] 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[169,173]. Targeted spatial transcriptomic methods can also be used in conjunction with scRNA-seq, where genes of interest that have already been identified can then be located more precisely[52,63]. In addition, probes for non-polyadenylated transcripts can be designed to query for other RNAs such as mature microRNAs and tRNAs[173].

Sequence information
In NGS-based and ISS-based methods, the cDNA sequence itself is a read out, enabling the detection of fusion transcripts[174], splice isoforms[58,175] and single-nucleotide variants and point mutations[60]. When integrated with the gene-expression matrix, these data can assist with reconstructing a time course, using RNA velocity[54] or lineage tracing[176].

Sensitivity
ISH-based methods are highly sensitive, recently reaching 80% detection efficiency relative to the gold-standard single-molecule fluorescence ISH (smFISH)[173]. The sensitivity of the NGS-based methods is significantly lower and remains inferior to scRNA-seq, but is rapidly improving to about 100 unique transcripts[55,58,59,177] per μm^2. There is generally a trade-off between sensitivity and gene throughput, as seen in the higher sensitivity of targeted ISS-based methods[64] relative to the unbiased methods[70].

Resolution
The resolution of in situ methods is limited only by the optical diffraction limit and, with expansion microscopy, has reached[80,173] around 100 nm. These methods are therefore well-suited to questions concerning sub-cellular organization. NGS-based methods are limited by the diameters of spots, but their resolution has rapidly increased since the original method[41], recently reaching[58,59] approximately 1 µm.

Area size
The in situ methods can span a wide range of sizes, although there is a trade-off between tissue size and imaging time[73]. By contrast, the NGS-based methods are standardized, with arrays on the order of 10 mm^2 (currently 6 mm^2 for the commercially available 10X Genomics Visium[49]), 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 of NGS-based methods. Commercialization has facilitated access in some cases, as seen with 10X Genomics Visium[49].

of tissue that can be assayed (Box 1), the methods reviewed here focus on technologies that enable 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 (ISS)-based methods—in which transcripts are amplified and sequenced in the tissue—and ISH-based methods—in which imaging probes are sequentially hybridized in the tissue[36–40] (Fig. 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 on a gene-expression matrix (Fig. 1d) that captures the transcriptome at every spot (that is, a pixel, a cell, or a group of cells).

Next-generation sequencing-based approaches

NGS-based approaches build on the conceptual innovations of scRNA-seq methodologies and are contingent on the addition of a spatial barcode before library preparation[35] (Fig. 1a). In 2016, Stahl et al. reported the first NGS-based method for spatial transcriptomics that enabled the capture of whole transcriptomes from tissue sections[41]. The central innovation was the capture of poly-adenylated RNA on spatially barcoded microarray slides before 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 1,000 spots (100 µm spot diameter with 200 µm centre-to-centre distance), 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 bulb[41], and has since been used by several other groups[44–48]. Visium[49], recently released by 10X Genomics, is an improved version of the technology with increased resolution (55 µm spot diameter with 100 µm centre-to-centre distance) and sensitivity (more than 10,000 transcripts per spot). This technology has been adopted in many different fields, including neuroscience[50], cancer biology[47,51] and developmental biology[52].

Slide-seq, another NGS-based technology, uses randomly barcoded beads deposited onto a slide for mRNA capture[53]. The position of each random barcode is obtained by in situ indexing. This method has achieved high resolution (10 µm) and sensitivity (500 transcripts per bead)[54]. In parallel, high-definition spatial transcriptomics (HDST) also improved the resolution, by replacing the glass slide with beads deposited in wells, similar to Slide-seq[55]. More recently, deterministic barcoding in tissue for spatial 'omics sequencing[56] (DBiT-seq) has adopted microfluidics to apply poly-T barcodes to the tissue section, whereas spatio-temporal enhanced resolution 'omics sequencing (Stereo-seq) uses randomly barcoded DNA nanoballs deposited in an array pattern to achieve nanoscale resolution[57]. Seq-scope has achieved subcellular resolution spatial barcoding and can be used to visualize nuclear and cytoplasmic transcripts[58]. An innovative approach was adopted in polony (or DNA cluster)-indexed library sequencing (PIXEL-seq), resulting in an increase in resolution of up to about 200-fold compared with existing methods[59].

In 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 approach to spatial transcriptomics have been introduced: ISS- and ISH-based methods. 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 (Fig. 1b). Using targeted probes for the reverse transcription, followed by sequencing-by-ligation[60], this method

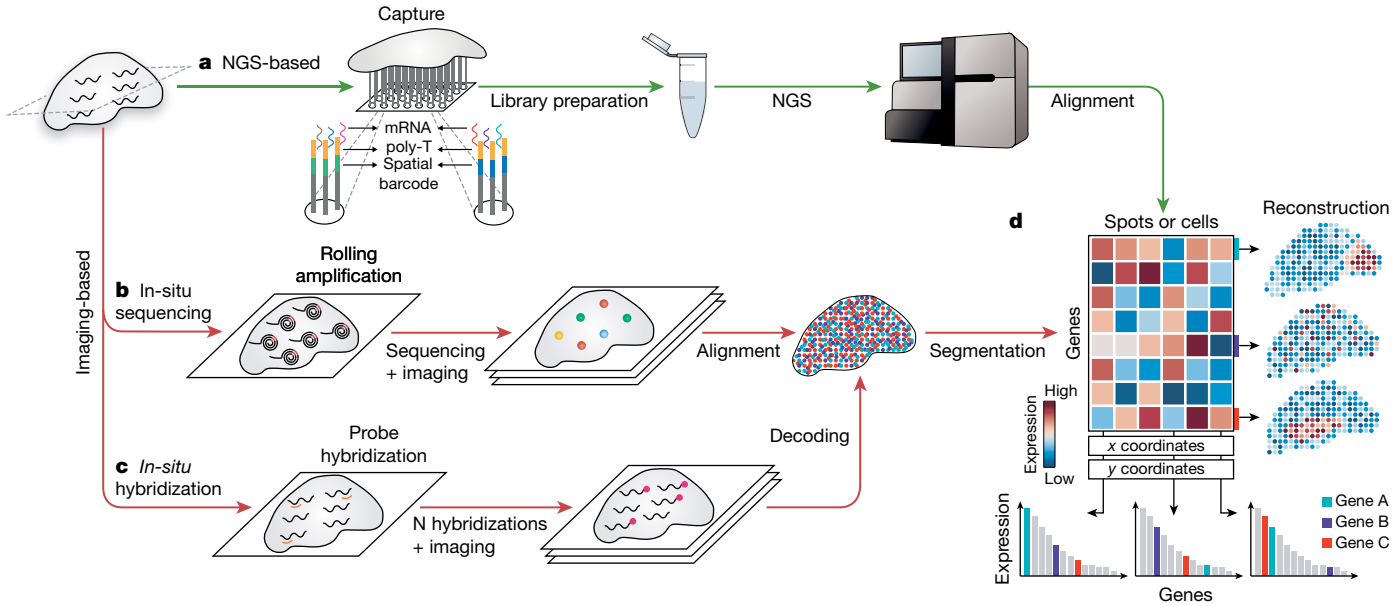

Fig. 1 | The technologies of spatial transcriptomics provide a gene-expression matrix. a, NGS-based spatial transcriptomic methods barcode transcripts according to their location in a lattice of spots. **b**, ISS approaches directly read out the transcript sequence within the tissue. **c**, ISH methods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

was implemented to study around 50 targeted genes in cancer[60,61], tuberculosis[62] and brain development[63]. Building on this approach, spatially resolved transcript amplicon readout mapping (STARMap) incorporated advances in hydrogel chemistry, improved padlock-probe and primer design and included an error-robust sequencing-by-ligation method, and was thus able to profile thousands of genes in the mouse cortex[64]. Other methods using sequencing-by-synthesis—for example, barcode in situ targeted sequencing (BaristaSeq)[65] and barcoded anatomy resolved by sequencing (Barseq)[66]—or sequencing by hybridization—for example, hybridization-based ISS (HybISS)[67]—have led to increased read lengths, enabling higher throughput and cellular barcoding. Furthermore, ISS has been combined with cDNA extraction for NGS[68,69], highlighting the difficulty in classifying spatial transcriptomic methods as either NGS- or imaging-based. ISS also has the potential for untargeted profiling, as demonstrated by fluorescent ISS (FISSEQ)[70]. Although the untargeted amplification can lead to optical crowding and lower sensitivity, the recently developed expansion sequencing (ExSeq) has demonstrated that expansion microscopy can be used to perform untargeted ISS in tissues[69].

ISH-based methods are the second category of imaging-based methods, and build on ISH technologies, whereby a target sequence is detected by hybridization of a complementary fluorescent probe (Fig. 1c). Initially limited in the number of distinguishable transcripts, innovations enabling the addition of sequential rounds of hybridization and imaging[71] combined with barcoding have enabled substantial multiplexing. In multiplexed error-robust fluorescence ISH (MERFISH), successive rounds of hybridizations are imaged to detect the presence or absence of fluorescently labelled probes. The serial images are then decoded, using the error-robust barcode associated with each transcript identity[72–74]. MERFISH has been used over a wide range of scales, from transcript location within individual cells[75] to tissue-level spatial transcriptomics, such as on the hypothalamic preoptic region[76]. Another strategy to increase the number of distinguishable transcripts is the combination of colours into pseudocolors, as done in sequential fluorescence ISH (seqFISH)[77,78]. Similar to MERFISH, this method can be applied to investigate intracellular organization[79] as well as to generate large maps—for example, of the hippocampus[78]. Both methods have improved considerably in the past few years, and are now able to detect around 10,000 genes at sub-cellular resolution[75,80]. Ongoing efforts in the community aim to improve the sensitivity and scale of these methods[34,36,38].

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 density[81]. 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 cells[82–84].

Insights into development, physiology and disease

As 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: spatial transcriptomics-based approaches have established detailed maps of the entire mouse brain[50], or of specific regions such as visual cortex[64], primary motor cortex[85], middle temporal gyrus[67], hypothalamic pre-optic region[76], hippocampus[69,78] and cerebellum[86]. Maynard et al. identified spatial patterns of known schizophrenia- and autism-related genes in their analysis of the dorsolateral prefrontal cortex, which led to proposed mechanisms of genetic susceptibility to schizophrenia[87].

In developmental biology, time-resolved spatial transcriptomics atlases have been useful to elucidate the spatial dynamics of heart development[52], spermatogenesis[88] and intestinal development[89]. 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 signalling in regulating differentiation towards ciliated or secretory epithelial cells[90]. These atlases have been the focus of coordinated community efforts to serve as effective resources for the research community, and are supported by the Human Cell Atlas project[91] and the Allen Institute for Brain Science[92].

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[44,51,93–96]. With the increasing recognition of the importance of the tumour microenvironment, spatial transcriptomics has been used to address its relationship to cancer cells adopting different states[45,46,69,95]. In particular, spatial transcriptomics enables the study of the molecular features across the boundaries between cancer and normal tissue. For example, an immunomodulatory cancer cell state was revealed in skin squamous cell carcinoma[47]. Spatial transcriptomics has also provided insights into the mechanisms of tissue dysregulation in neurodegenerative disorders including Alzheimer's disease[97,98] and amyotrophic lateral sclerosis[99], infectious and inflammatory processes such as leprosy[100], influenza[101] and sepsis[102], and rheumatological diseases including rheumatoid arthritis and spondyloarthritis[103,104].

Exploratory data analysis

Spatial transcriptomic technologies result in a gene-expression matrix, which can be analysed 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 Tukey[105]—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 and 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[106,107].

Analysing 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 Giotto[108], Seurat[109,110], STutility[111] and stLearn[112]. 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 artefact, 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 as 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 neighbouring spots, averaging gene expression between physically adjacent spots in a moving window along the spatial coordinates can reduce noise[47]. To compare the expression of a gene across spots, transcriptomes are often normalized by dividing by the total number of transcripts (transcripts per million (TPM)) or using regularized negative binomial regression[113]. 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 (Fig. 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 operation that have been used to study spatial transcriptomic data, although more operations will undoubtedly be devised (Fig. 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 (Fig. 2b).

Cluster

The clustering operation reveals structure in the data, which, at its most basic defines sets of spots with similar transcriptomes, or orthogonally, identifies 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 principal component analysis (PCA), *t*-distributed stochastic neighbour embedding (*t*-SNE)[114] and uniform manifold approximation and projection (UMAP)[115]. These similarities are then used to cluster spots—for example, using k-means, Louvain or hierarchical clustering[116]. 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 inflammatory[117]. 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 petals[118].

Gene clustering, using the same approach, can identify co-expressed gene modules corresponding to a cell type or cell state[108]. In spatial transcriptomic data from the cerebellum, for example, clustering of genes identified two modules of spatially correlated genes in Purkinje cells[53]. Methods to cluster genes and spots simultaneously have also been used, including non-negative matrix factorization[119,120] or factor analysis[93], where the gene-expression matrix is factorized to reveal the underlying structure in spot clusters and gene modules. In prostate tumour samples, this revealed sets of spots and genes corresponding to cancer, stroma and inflammation[93]. Clustering methods such as BayesSpace[121], that focus on the specific features of spatial transcriptomics are currently being developed.

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[53,54], or the interface between tumour and microenvironment[85,122]. Orthogonally, the analysis may be focused on context-specific genes, either chosen a priori from biological knowledge—most notably in imaging-based methods that do not yet cover the whole transcriptome—or chosen from the dataset itself, for example by identifying highly variable genes. 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*)[123], neighbour enrichment (for example, in BinSpect)[108] or entropy (for example, in Haystack)[124]. Trendsceek[125] uses a marked point processes approach[126] and is able to identify hotspots, streaks and gradients of expression. SpatialDE decomposes a given gene's expression variability into spatial and non-spatial components using Gaussian process regression[127], and a similar approach was extended upon in SPARK[128]. Cancer-specific metabolic vulnerabilities were characterized by identifying spatially variable genes in prostate cancer using this approach[94].

Score

While the genes and spots are the primary data observations of spatial transcriptomics, the underlying biology means 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 the data in this way can identify

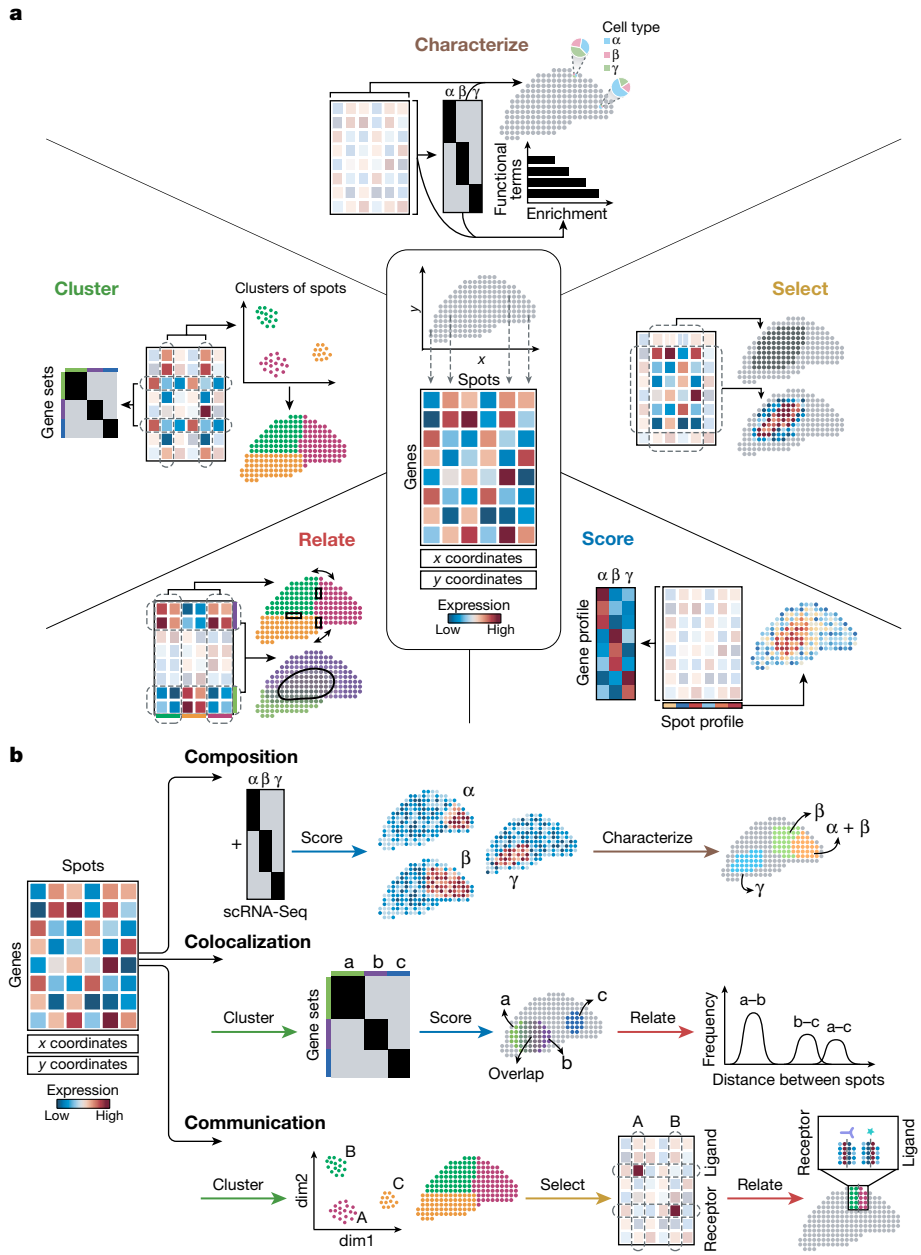

Fig. 2 | Exploratory data analysis using spatial transcriptomic datasets. **a**, Schematic of exploratory data analysis operations of spatial transcriptomic datasets. Characterize: depicted are spots characterized to be composed of proportions of cell type α, β and γ and gene sets annotated with functional 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 on the basis of 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, and 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. Dim, dimension.

functional properties—for example, a stress-response state or infiltrating macrophages that are spatially organized within a tumour—which might not be detectable when analysing 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 workflow[110]. In the brain, for example, Moffitt et al. generated average cell-type expression profiles to compare spatial transcriptomics and scRNA-seq clusters[76]. In melanoma, spots were scored according to their expression of previously established gene sets corresponding to cancer cell states[45] or to Gene Ontology terms[122].

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. Integration with other data sources and with other prior knowledge are essential to achieve this. 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 brain[76], and in pancreatic cancer samples to annotate normal and malignant regions of the tumour[46]. 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[46] (MIA) and of gene-set enrichment analysis[129] (GSEA), which queries for enrichment with functional groups obtained from Gene Ontology, Kyoto Encyclopedia of Genes and Genomes (KEGG), Hallmark and other databases[129–131].

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[52,90]. Most methods achieve this by integrating single-cell data, generated either from the same sample (paired) or from a similar sample or database (unpaired). This integration helps to overcome the limitations of scRNA-seq—which lacks spatial information—and NGS-based spatial transcriptomics—which is not at single-cell resolution. By performing non-negative least squares (NNLS) linear regression on the spatial transcriptomic data using the NMF factors derived from single-cell data, cell-type composition of spots can be inferred, as implemented in SPOTlight[132] and NMFreg[53]. Methods tailored to cell-type decomposition of spatial transcriptomic data have been devised, including probability-based methods[133–137], graph-based methods[138], and deep learning methods[139,140]. In Stereoscope[133], the cell-type parameters are assigned by maximum-likelihood estimation on the single cell and are used to estimate the cell-type composition of each spot. Cell2location[134] additionally attempts to infer the absolute number of cells per spot. Deconvoluting spatial transcriptomics data through graph-based artificial intelligence (DSTG)[138] uses single-cell data to construct pseudospots, and then links real and pseudospots in a graph of nearest neighbours. The spatial dampened weighted least-squares (spatialDWLS) method borrows from methodologies previously used for bulk RNA-seq deconvolution and applies cell-type enrichment followed by a dampened weighted least-squares method to determine spot composition[141].

Spatial transcriptomics methods with sub-cellular resolution face the inverse problem of grouping spots into organelles or cells. Seq-scope uses transcript annotations as spliced, unspliced or mitochondrial to identify the cytoplasm, nucleus and mitochondria within cells[58]. Recent approaches have been developed that use the local density of each RNA species to assign a cell type to each spot[82,83]. Probabilistic cell typing by ISS (pciSeq) is able to identify cell types more efficiently in larger tissue areas[23,84]. Fluorescence ISH–iterative cell-type assignment (FICT) integrates expression and neighbourhood information to assign cell types[142]. 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 RNAs[84].

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, or developmental or functional relationships. For example, Stickels et al.[54] 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 have been devised using methods that were originally developed for scRNA-seq. RNA velocity[143] 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 neural development[54]. RNA-seq-based copy-number variation inference identifies chromosomal aneuploidies, which can be used to distinguish malignant spots from non-malignant spots and to identify distinct subclones[144,145]. When two sets of spots are spatially adjacent, potential modes of interaction[146] between the cells can be proposed by examining their paired receptors and ligands[108] using known databases such as CellPhoneDB[47,90,147] or NicheNet[148].

Hypothesis generation and testing

Spatial transcriptomics atlases of healthy or diseased tissues naturally lend themselves to unbiased exploration and hypothesis generation[52,90]. Even those 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 hypotheses[98] (Fig. 3). Thousands of spots or genes can 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 immunofluorescence[46] or ISH[76] (Fig. 3a).

Alternatively, spatial transcriptomic data can be incorporated into classical hypothesis-driven experimental design, whereby a sufficiently powered experiment is used 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 data point, to be compared across replicates and conditions, necessitating that data be collected in sufficient numbers to ensure statistical rigour 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 be further tested in model systems, in vitro or in vivo, or with clinical data (Fig. 3b).

Integration 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. While currently often 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 regions[2]. 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 composition[112]. Another study improved the resolution of spatial transcriptomics gene-expression data by fusing it with high-resolution histology image data[149]. Deep learning has also been used to predict cell-type annotations from gene expression and histology, outperforming annotations predicted from either modality alone[150–152]. With the increase in transcriptomic data available for training, machine learning algorithms have also been used to predict gene expression from histopathology images[153,154]. 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

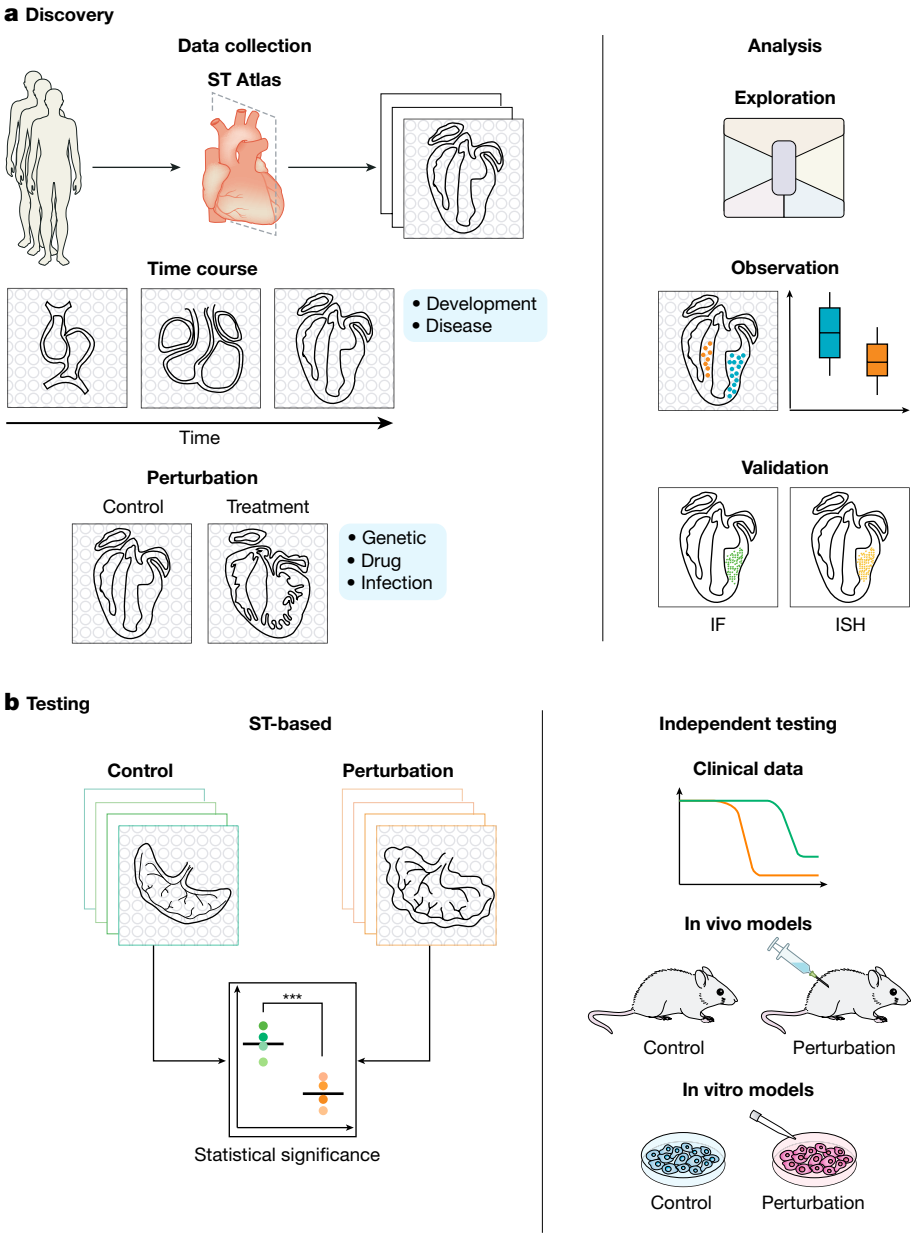

Fig. 3 | Hypothesis generation and testing using spatial transcriptomics. a, Spatial transcriptomics (ST) can be used for hypothesis generation in various experimental contexts. Examples of spatial transcriptomic datasets include normal tissue (ST atlas), a developmental or disease time course, and perturbation 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.

nuclear bodies and zone interfaces in embryonic stem cells[155]. Integrating spatial transcriptomic datasets with high-throughput imaging of the genome in situ and the spatial distribution of histone marks within a tissue will be extremely valuable[156–158]. Recently, spatial mapping of genome organization with concurrent DNA sequencing within intact tissue has been made feasible[159]. This suggests that the goal of combining spatial genome sequencing with in situ transcriptomic profiling may be within reach, deepening our understanding of how genome organization and function are encoded[158].

Augmenting gene-expression data with a complementary modality such as protein co-detection can also shed light onto processes that are not captured by spatial transcriptomics, such as post-translational modification and sub-cellular localization of proteins and their dysregulation in disease. Targeted protein co-detection can be performed alongside spatial transcriptomics using immunostaining on the same tissue section, as enabled by Visium[49]. A novel imaging cytometry-based approach has also been used to simultaneously detect transcripts and proteins in breast cancer tissue samples[160].DBiT-seq enables the co-mapping of mRNA and proteins in the tissue using antibody-derived DNA tags, as is done in cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq)[161]. High-throughput spatial methods for protein detection such as multiplexed ion beam imaging (MIBI), co-detection by indexing (CODEX), tissue cyclic immunofluorescence (t-cyCIF) and automated mass spectrometry provide an unparalleled snapshot of the proteome within the tissue section[162–167]. Technological advances that allow the integration of these high-throughput proteomics methods with spatial transcriptomics will greatly 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[49,70,168,169]. 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 changes in gene expression 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 in the field, the Human Genome Project may serve as a useful parallel. The initial draft of the human genome was published in 2001[170,171] and provided a reference for studying 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 be only the first step towards elucidating organizing principles of tissue biology. It is the coupling of these high-resolution cellular atlases with hypothesis-free inquiries that will enable new insights 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 neighbours—was sufficient to recover complex spatial patterns in the *Drosophila* embryo[172]. Building on this idea, the exploration of spatial transcriptomic datasets will enable us to uncover the fundamental principles that guide our modelling 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03634-9.

1. Barresi, M. J. F. & Gilbert, S. F. *Developmental Biology* (Sinauer Associates, 2019).
2. Damjanov, I. & McCue, P. A. *Histopathology: A Color Atlas and Textbook* (Lippincott Williams & Wilkins, 1996).
3. Safai, B. & Good, R. A. *Immunodermatology* (Springer Science & Business Media, 2013).
4. Lehmann, R. & Tautz, D. in *Methods in Cell Biology* Vol. 44 (eds Lawrence, S. B. & Fyrberg, E. A.) 575–598 (Academic Press, 1994).
5. Swanson, P. E. Foundations of immunohistochemistry. A practical review. *Am. J. Clin. Pathol.* **90**, 333–339 (1988).
6. Mincarelli, L., Lister, A., Lipscombe, J. & Macaulay, I. C. Defining cell identity with single-cell omics. *Proteomics* **18**, e1700312 (2018).
7. Macaulay, I. C., Ponting, C. P. & Voet, T. Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* **33**, 155–168 (2017).
8. Tanay, A & Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
9. Xia, B. & Yanai, I. A periodic table of cell types. *Development* **146**, dev169854 (2019).
10. Stubbington, M. J. T., Rozenblatt-Rosen, O., Regev, A. & Teichmann, S. Single-cell transcriptomics to explore the immune system in health and disease. Science **358**, 58–63 (2017).
11. Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N. & Werb, Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nat. Cell Biol.* **20**, 1349–1360 (2018).
12. Papalexi, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2018).
13. Combs, P. A. & Eisen, M. B. Sequencing mRNA from cryo-sliced *Drosophila* embryos to determine genome-wide spatial patterns of gene expression. *PLoS ONE* **8**, e71820 (2013).
14. Junker, J. P. et al. Genome-wide RNA tomography in the zebrafish embryo. *Cell* **159**, 662–675 (2014).
15. Lacraz, G. P. A. et al. Tomo-seq identifies SOX9 as a key regulator of cardiac fibrosis during ischemic injury. *Circulation* **136**, 1396–1409 (2017).
16. van den Brink, S. C. et al. Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* **582**, 405–409 (2020).
17. Nichterwitz, S. et al. Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat. Commun.* **7**, 12139 (2016).
18. Nichterwitz, S., Benitez, J. A., 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 Mol. Biol.* **1649**, 95–110 (2018).
19. Aguila, J. et al. Spatial transcriptomics and in silico random pooling identify novel markers of vulnerable and resistant midbrain dopamine neurons. Preprint at https://doi.org/10.1101/334417 (2021).
20. Moor, A. E. et al. Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. *Cell* **175**, 1156–1167.e15 (2018).
21. Chen, J. et al. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat. Protoc.* **12**, 566–580 (2017).
22. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
23. Peng, G. et al. Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* **572**, 528–532 (2019).
24. Geiss, G. K. et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 (2008).
25. Halpern, K. B. et al. Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. *Nat. Biotechnol.* **36**, 962–970 (2018).
26. Boisset, J.-C. et al. Mapping the physical network of cellular interactions. *Nat. Methods* **15**, 547–553 (2018).
27. Giladi, A. et al. Dissecting cellular crosstalk by sequencing physically interacting cells. *Nat. Biotechnol.* **38**, 629–637 (2020).
28. Manco, R. et al. Clump sequencing exposes the spatial expression programs of intestinal secretory cells. *Nat. Commun.* **12**, 3074 (2021).
29. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
30. Achim, K. et al. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
31. Pettit, J.-B. et al. Identifying cell types from spatially referenced single-cell expression datasets. *PLOS Comput. Biol.* **10**, e1003824 (2014).
32. Karaiskos, N. et al. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017).
33. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).
34. Zhuang, X. Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat. Methods* **18**, 18–22 (2021).
35. Larsson, L., Frisén, J. & Lundeberg, J. Spatially resolved transcriptomics adds a new dimension to genomics. *Nat. Methods* **18**, 15–18 (2021).
36. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015).
37. Moor, A. E. & Itzkovitz, S. Spatial transcriptomics: paving the way for tissue-level systems biology. *Curr. Opin. Biotechnol.* **46**, 126–133 (2017).
38. Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially resolved transcriptomes-next generation tools for tissue exploration. *BioEssays* **42**, e1900221 (2020).
39. Waylen, L. N., Nim, H. T., Martelotto, L. G. & Ramialison, M. From whole-mount to single-cell spatial assessment of gene expression in 3D. *Commun. Biol.* **3**, 602 (2020).
40. Teves, J. M. & Won, K. J. Mapping cellular coordinates through advances in spatial transcriptomics technology. *Mol. Cells* **43**, 591–599 (2020).
41. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
This paper was the first to perform array-based spatial transcriptomics, using positional barcodes at a resolution of 200 µm, 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).
43. Salmén, F. et al. Barcoded solid-phase RNA capture for spatial transcriptomics profiling in mammalian tissue sections. *Nat. Protoc.* **13**, 2501–2534 (2018).
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).
45. Baron, M. et al. The stress-like cancer cell state is a consistent component of tumorigenesis. *Cell Syst.* **11**, 536–546.e7 (2020).
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).
47. Ji, A. L. et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 497–514.e22 (2020).

48. Hildebrandt, F. et al. Spatial transcriptomics to define transcriptional patterns of zonation and structural components in the liver. Preprint at https://doi.org/10.1101/2021.01.11.426100 (2021).
49. Spatial Transcriptomics. *10x Genomics* https://www.10xgenomics.com/spatial-transcriptomics/ (2021).
50. Ortiz, C. et al. Molecular atlas of the adult mouse brain. *Sci. Adv.* **6**, eabb3446 (2020).
51. Grauel, A. L. et al. TGFβ-blockade uncovers stromal plasticity in tumors by revealing the existence of a subset of interferon-licensed fibroblasts. *Nat. Commun.* **11**, 6315 (2020).
52. Asp, M. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660.e19 (2019).
53. Rodriques, S. G. et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
This paper describes Slide-seq, an array-based method with 10μm resolution, performed in the cerebellum and hippocampus.
54. Stickels, R. R. et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
55. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
56. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681.e18 (2020).
57. Chen, A. et al. Large field of view-spatially resolved transcriptomics at nanoscale resolution. Preprint at https://doi.org/10.1101/2021.01.17.427004 (2021).
58. Cho, C.-S. et al. Microscopic examination of spatial transcriptome using Seq-scope. *Cell* **184**, 3559–3572.e22 (2021).
59. Fu, X. et al. Continuous polony gels for tissue mapping with high resolution and RNA capture efficiency. Preprint at https://doi.org/10.1101/2021.03.17.435795 (2021).
60. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
This was the first report of ISS, which was used to map the expression of 31 transcripts using four-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).
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).
63. Tiklová, K. et al. Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nat. Commun.* **10**, 581 (2019).
64. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
65. Chen, X., Sun, Y.-C., Church, G. M., Lee, J. H. & Zador, A. M. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res.* **46**, e22 (2018).
66. Chen, X. et al. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179**, 772–786.e19 (2019).
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).
68. Fürth, D., Hatini, V. & Lee, J. H. In situ transcriptome accessibility sequencing (INSTA-seq). Preprint at https://doi.org/10.1101/722819 (2019).
69. Alon, S. et al. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, eaax2656 (2021).
70. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
This paper introduced an untargeted ISS method, FISSEQ, that generated 30 base reads from 8,102 genes in human primary fibroblasts.
71. Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
72. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
73. Moffitt, J. R. et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl Acad. Sci. USA* **113**, 11046–11051 (2016).
74. Wang, G., Moffitt, J. R. & Zhuang, X. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep.* **8**, 4847 (2018).
75. Xia, C., Babcock, H. P., Moffitt, J. R. & Zhuang, X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Rep.* **9**, 7721 (2019).
76. Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
77. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
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).
79. Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376.e16 (2018).
80. Eng, C. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
81. Najman, L. & Schmitt, M. Watershed of a continuous function. *Signal Processing* **38**, 99–112 (1994).
82. Park, J. et al. Segmentation-free inference of cell types from in situ transcriptomics data. Preprint at https://doi.org/10.1101/800748 (2020).
83. Littman, R. et al. JSTA: joint cell segmentation and cell type annotation for spatial transcriptomics. Preprint at https://doi.org/10.1101/2020.09.18.304147 (2020).
84. Qian, X. et al. Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* **17**, 101–106 (2020).
85. BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. Preprint at https://doi.org/10.1101/2020.10.19.343129.
86. Kebschull, J. M. et al. Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set. *Science* **370**, eabd5059 (2020).
87. Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–1436 (2021).
88. Chen, H. et al. Dissecting mammalian spermatogenesis using spatial transcriptomics. Preprint at https://doi.org/10.1101/2020.10.17.343335 (2020).
89. Fawkner-Corbett, D. et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826.e23 (2021).
90. Garcia-Alonso, L. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. Preprint at https://doi.org/10.1101/2021.01.02.425073 (2021).
91. 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).
92. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
93. Berglund, E. et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* **9**, 2419 (2018).
94. 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).
95. Hwang, W. L. et al. Single-nucleus and spatial transcriptomics of archival pancreatic cancer reveals multi-compartment reprogramming after neoadjuvant treatment. https://doi.org/10.1101/2020.08.25.267336 (2020).
96. Smith, E. A. & Hodges, H. C. The spatial and genomic hierarchy of tumor ecosystems revealed by single-cell technologies. *Trends Cancer* **5**, 411–425 (2019).
97. Navarro, J. F. et al. Spatial transcriptomics reveals genes associated with dysregulated mitochondrial functions and stress signaling in Alzheimer disease. *iScience* **23**, 101556 (2020).
98. Chen, W.-T. et al. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* **182**, 976–991.e19 (2020).
99. Maniatis, S. et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89–93 (2019).
100. Ma, F. et al. Single cell and spatial transcriptomics defines the cellular architecture of the antimicrobial response network in human leprosy granulomas. https://doi.org/10.1101/2020.12.01.406819 (2020).
101. Boyd, D. F. et al. Exuberant fibroblast activity compromises lung function via ADAMTS4. *Nature* **587**, 466–471 (2020).
102. Janosevic, D. et al. The orchestrated cellular and molecular responses of the kidney to endotoxin define a precise sepsis timeline. *eLife* **10**, e62270 (2021).
103. Carlberg, K. et al. Exploring inflammatory signatures in arthritic joint biopsies with Spatial Transcriptomics. *Sci. Rep.* **9**, 18975 (2019).
104. Vickovic, S. et al. Three-dimensional spatial transcriptomics uncovers cell type dynamics in the rheumatoid arthritis synovium. Preprint at https://doi.org/10.1101/2020.12.10.420463 (2020).
105. Tukey, J. W. *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.
106. Yanai, I. & Lercher, M. What is the question? *Genome Biol.* **20**, 289 (2019).
107. Yanai, I. & Lercher, M. The data-hypothesis conversation. *Genome Biol.* **22**, 58 (2021).
108. Dries, R. et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).
109. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
110. 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).
111. Bergenstråhle, J., Larsson, L. & Lundeberg, J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 482 (2020).
112. Pham, D. et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. Preprint at https://doi.org/10.1101/2020.05.31.125658 (2020).
113. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
114. Zhou, B. & Jin, W. Visualization of single cell RNA-seq data using *t*-SNE in R. *Methods Mol. Biol.* **2117**, 159–167 (2020).
115. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018).
116. Kiselev, V. Y., Andrews, T. S. & Hemberg, M. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* **20**, 273–282 (2019).
117. Lundmark, A. et al. Gene expression profiling of periodontitis-affected gingival tissue by spatial transcriptomics. *Sci. Rep.* **8**, 9370 (2018).
118. Giacomello, S. et al. Spatially resolved transcriptome profiling in model plant species. *Nat. Plants* **3**, 17061 (2017).
119. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
120. Gao, Y. & Church, G. Improving molecular cancer class discovery through sparse non-negative matrix factorization. *Bioinformatics* **21**, 3970–3975 (2005).
121. Zhao, E. et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00935-2 (2021).
122. Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I. & White, R. M. Spatial transcriptomics reveals the architecture of the tumor/microenvironment interface. Preprint at https://doi.org/10.1101/2020.11.05.368753 (2021).
123. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
124. Vandenbon, A. & Diez, D. A clustering-independent method for finding differentially expressed genes in single-cell transcriptome data. *Nat. Commun.* **11**, 4318 (2020).
125. Edsgärd, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods* **15**, 339–342 (2018).
126. Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. *Statistical Analysis and Modelling of Spatial Point Patterns* (John Wiley & Sons, 2008).
127. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).

128. Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* **17**, 193–200 (2020).
129. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
130. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
131. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes And Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
132. Elosua, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: Seeded NMF regression to Deconvolute Spatial Transcriptomics Spots with Single-Cell Transcriptomes. https://doi.org/10.1101/2020.06.03.131334.
133. Andersson, A. et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* **3**, 565 (2020).
134. Kleshchevnikov, V. et al. Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. Preprint at https://doi.org/10.1101/2020.11.15.378125 (2020).
135. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00830-w (2021).
136. Lopez, R. et al. Multi-resolution deconvolution of spatial transcriptomics data reveals continuous patterns of inflammation. Preprint at https://doi.org/10.1101/2021.05.10.443517 (2021).
137. Zeira, R., Land, M. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. Preprint at https://doi.org/10.1101/2021.03.16.435604 (2021).
138. Su, J. & Song, Q. DSTG: Deconvoluting Spatial Transcriptomics Data through Graph-based artificial intelligence. *Brief. Bioinform.* https://doi.org/10.1093/bib/bbaa414 (2021).
139. Biancalani, T. et al. Deep learning and alignment of spatially-resolved whole transcriptomes of single cells in the mouse brain with Tangram. Preprint at https://doi.org/10.1101/2020.08.29.272831 (2020).
140. Nelson, M. E., Riva, S. G. & Cvejic, A. SMaSH: A scalable, general marker gene identification framework for single-cell RNA sequencing and spatial transcriptomics. Preprint at https://doi.org/10.1101/2021.04.08.438978 (2021).
141. Dong, R. & Yuan, G.-C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. Preprint at https://doi.org/10.1101/2021.02.02.429429 (2021).
142. Teng, H., Yuan, Y. & Bar-Joseph, Z. Cell type assignments for spatial transcriptomics data. Preprint at https://doi.org/10.1101/2021.02.25.432887 (2021).
143. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
144. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
145. Elyanow, R., Zeira, R., Land, M. & Raphael, B. STARCH: Copy number and clone inference from spatial transcriptomics data. *Phys. Biol.* **18**, 035001 (2021).
146. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).
147. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
148. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
149. Bergenstråhle, L. et al. Super-resolved spatial transcriptomics by deep data fusion. Preprint at https://doi.org/10.1101/2020.02.28.963413 (2020).
150. Tan, X., Su, A., Tran, M. & Nguyen, Q. SpaCell: integrating tissue morphology and spatial gene expression to predict disease cells. *Bioinformatics* **36**, 2293–2294 (2020).
151. Monjo, T., Koido, M., Nagasawa, S., Suzuki, Y. & Kamatani, Y. Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. Preprint at https://doi.org/10.1101/2021.04.22.440763 (2021).
152. Bao, F. et al. Characterizing tissue composition through combined analysis of single-cell morphologies and transcriptional states. Preprint at https://doi.org/10.1101/2020.09.05.284539 (2021).
153. He, B. et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat. Biomed. Eng.* **4**, 827–834 (2020).
154. Levy-Jurgenson, A., Tekpli, X., Kristensen, V. N. & 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).
155. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
156. Deng, Y. et al. Spatial epigenome sequencing at tissue scale and cellular level. Preprint at https://doi.org/10.1101/2021.03.11.434985 (2021).
157. Nguyen, H. Q. et al. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* **17**, 822–832 (2020).
158. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* **182**, 1641–1659.e26 (2020).
159. Payne, A. C. et al. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* **371**, eaay3446 (2021).
160. Schulz, D. et al. Simultaneous multiplexed imaging of mRNA and proteins with subcellular resolution in breast cancer tissue samples by mass cytometry. *Cell Syst.* **6**, 531 (2018).
161. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
162. Angelo, M. et al. Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* **20**, 436–442 (2014).
163. Keren, L. et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell* **174**, 1373–1387.e19 (2018).
164. Piehowski, P. D. et al. Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100-µm spatial resolution. *Nat. Commun.* **11**, 8 (2020).
165. Lin, J.-R. et al. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018).
166. Goltsev, Y. et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* **174**, 968–981.e15 (2018).
167. Kohman, R. E. & Church, G. M. Fluorescent in situ sequencing of DNA barcoded antibodies. Preprint at https://doi.org/10.1101/2020.04.27.060624 (2020).
168. Liu, Y., Enninful, A., Deng, Y. & Fan, R. Spatial transcriptome sequencing of FFPE tissues at cellular level. Preprint at https://doi.org/10.1101/2020.10.13.338475 (2020).
169. Nagarajan, M. B., Tentori, A. M., Zhang, W. C., Slack, F. J. & Doyle, P. S. Spatially resolved and multiplexed MicroRNA quantification from tissue using nanoliter well arrays. *Microsyst. Nanoeng.* **6**, 51 (2020).
170. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
171. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
172. Nitzan, M., Karaiskos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* **576**, 132–137 (2019).
173. 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. USA* **116**, 19490–19499 (2019).
174. Friedrich, S. & Sonnhammer, E. L. L. Fusion transcript detection using spatial transcriptomics. *BMC Med. Genomics* **13**, 110 (2020).
175. 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).
176. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
177. Rodriques, S. G. et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

Acknowledgements We thank F. Kuperwaser, A. Pountain, B. Xia and other members of the Yanai laboratory, as well as M. Phillips for critical reading and feedback. We thank the students of the exploratory data analysis course at NYU Langone Health. I.Y. was supported by grants from the NIH (R01AI143290 and R01LM013522) and the Lowenstein Foundation, and D.B. was supported by the National Institutes for Health (F30CA257400).

Author contributions A.R., D.B. and I.Y. collectively wrote the review. A.R., D.B., G.S.F. and I.Y. edited and revised the manuscript and conceptualized the figures, which G.S.F. then designed and created.

Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to I.Y.

Peer review information *Nature* thanks Guo-Cheng Yuan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.