nature biotechnology

REVIEW ARTICLE

https://doi.org/10.1038/s41587-022-01448-2



# The expanding vistas of spatial transcriptomics

Luyi Tian[1], Fei Chen[1,2] ✉ and Evan Z. Macosko[1,3] ✉

**The formation and maintenance of tissue integrity requires complex, coordinated activities by thousands of genes and their encoded products. Until recently, transcript levels could only be quantified for a few genes in tissues, but advances in DNA sequencing, oligonucleotide synthesis and fluorescence microscopy have enabled the invention of a suite of spatial transcriptomics technologies capable of measuring the expression of many, or all, genes in situ. These technologies have evolved rapidly in sensitivity, multiplexing and throughput. As such, they have enabled the determination of the cell-type architecture of tissues, the querying of cell–cell interactions and the monitoring of molecular interactions between tissue components. The rapidly evolving spatial genomics landscape will enable generalized high-throughput genomic measurements and perturbations to be performed in the context of tissues. These advances will empower hypothesis generation and biological discovery and bridge the worlds of tissue biology and genomics.**

New technologies to measure biomolecules have been critical drivers of biological progress. When measuring biomolecules, researchers have historically faced a key trade-off in selecting experimental methodologies. On the one hand, '-omics' tools generate broad, often comprehensive measurements of many biomolecules from a purified specimen. On the other hand, a suite of targeted tools, such as immunostaining or in situ hybridization, can localize a much smaller number of specific molecules within intact cells and tissues. Research projects, therefore, often have combined these two phases: a researcher first formulates a hypothesis using an 'omics technology and then performs targeted, hypothesis-driven work to characterize the role of specific genes or proteins within intact tissues of interest.

This historical bifurcation in methodologies is being upended by the recent rapid development of spatial transcriptomics (ST) technologies. These tools enable the quantification of RNAs across the transcriptome within intact tissue sections. Broadly speaking, we see ST as best suited to answering three kinds of biological questions (Fig. 1a). First, ST technologies can elucidate the cell-type composition of tissues. Cell-type definitions are frequently imported from large-scale single-cell RNA-seq or epigenetics datasets and computationally projected onto ST datasets to learn their spatial distributions, but definitions can also be generated from the ST data directly. To date, this has been the most used application of ST in published studies. Using ST, compositional atlases have been generated for a wide variety of tissues, including nervous system tissues[1–5], human kidney[6–8], heart[9], testes[10] and lung[11].

A second kind of question relates to cellular interactions; which rules and patterns define how individual cell types spatially covary with each other? For example, an ST study of mouse visual cortex[4] found a predilection for inhibitory neuron subtypes to be more spatially proximal to each other than would be expected if they were distributed randomly. Similar kinds of proximity analyses have been used to define gene expression patterns upregulated by amyloid plaques found in Alzheimer's disease[12] and to characterize histopathological responses to traumatic brain injury[13].

Finally, because ST technologies often deliver transcriptome-wide data on gene expression in situ, ST can help elucidate molecular interactions between tissue components. By defining ligand–receptor pairs among cell types at different spatial proximities, we can determine whether, and how, individual cell types are communicating with each other. Such analyses should help clarify many cell non-autonomous phenomena, including interactions between tumors and the surrounding environment[14,15], immune infiltrates in tissues or the establishment of developmental gradients[16].

To understand how current and future technologies can improve our capacity for addressing these problems, here we describe in detail the existing ST technologies in the sections below, including those that use either next-generation sequencing for gene detection or imaging-based detection. ST emerged from parallel, synchronous efforts by two distinct groups of technologists (Fig. 1b). In genomics, advances in massively parallel DNA sequencing, molecular biology, DNA-based molecular barcoding and computational analysis made possible the measurement of gene expression, and more recently epigenetic regulation, within many individual cells. These strategies and concepts were creatively adapted to capture RNA locally from intact tissue sections on a pixelated, DNA-barcoded surface and read out their gene identities using next-generation sequencing[17]. We term this family of technologies 'sequencing-based ST' (sST). In parallel, technologists working on microscopy techniques developed several strategies for simultaneously detecting the presence of many mRNA transcripts within tissues using fluorescence in situ hybridization (FISH)[18,19] or direct in situ sequencing[20,21]. We call such technologies 'imaging-based ST' (iST). These two classes of technology deliver similar and complementary measurements of gene expression in situ.

In our discussion of the major instantiations of sST and iST, we describe the key metrics and experimental frameworks for their characterization and validation. In particular, we emphasize the importance to the field of adopting experiments that validate technological claims and enable comparisons across platforms. The design of quality-controlled experiments with well-defined conditions that can be easily replicated across labs and methodologies can be extremely catalytic for technology development. For example, the mixed-species experiment used in single-cell technology development[22,23] (in which cells from well-established cell lines derived from different organisms are mixed together before the assay) has become a widely accepted benchmark for validating the sensitivity and specificity of new methodologies. The field of ST could similarly benefit from defining a standard set of reference

[1]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Harvard Stem Cell and Regenerative Biology, Cambridge, MA, USA. [3]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. ✉e-mail: chenf@broadinstitute.org; emacosko@broadinstitute.org

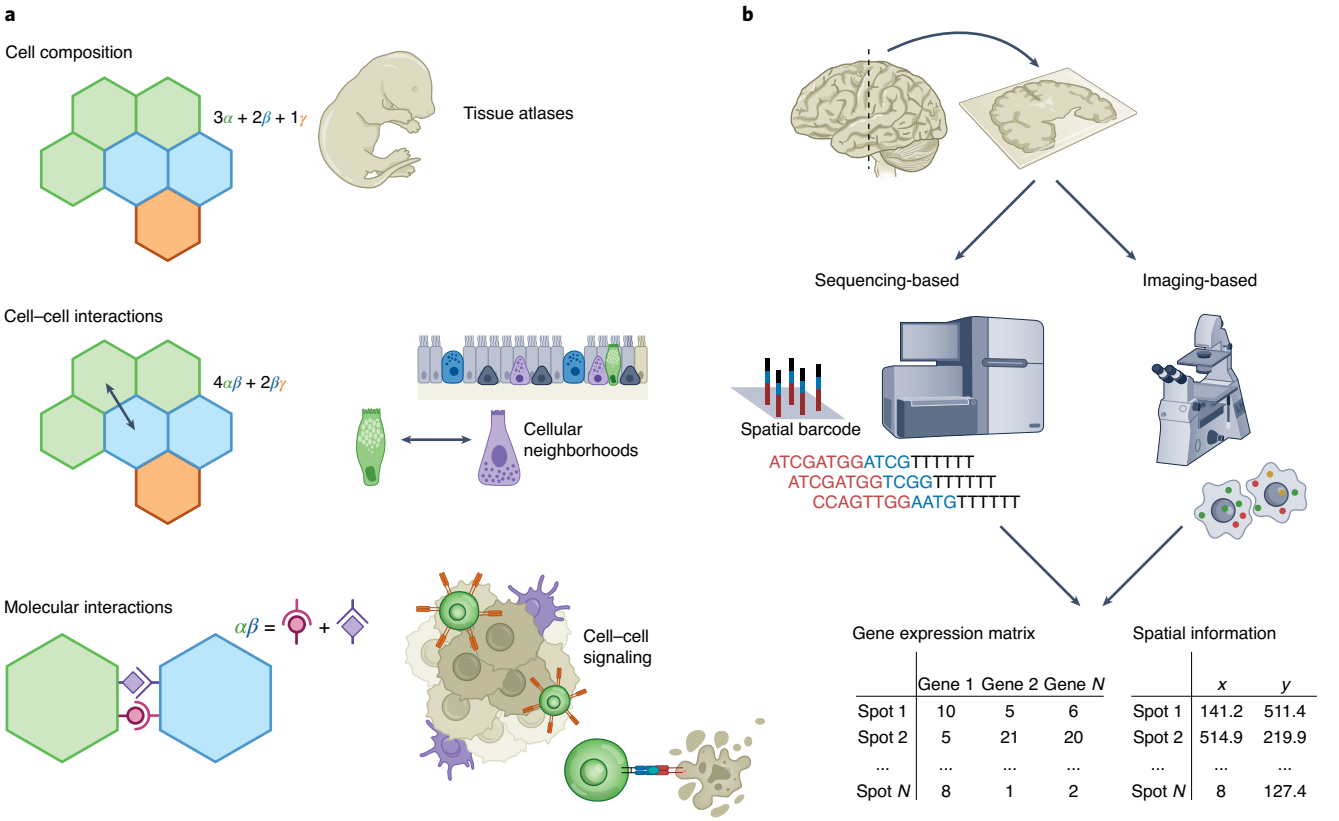


**Fig. 1 | Usage of ST in biological experimentation. a**, Three primary classes of biological questions addressed by ST. Colored hexagons represent different types of cells indicated by $\alpha$, $\beta$ and $\gamma$ in the corresponding model equations. **b**, Summary of the two classes of ST technology: sequencing-based and imaging-based methods. The input to ST technologies generally includes tissue sections, and the output is generally a digital gene expression matrix and a table of spatial coordinates for each spot. Detailed descriptions of individual methods within these two classes are shown in Figs. 2 and 3.

experiments by which to gauge the performance of the key experimental parameters.

## sST

sST technologies typically begin with the construction of a spatially indexed surface (Fig. 2a), in which each pixel contains a barcoded DNA primer that uniquely marks a pixel's location in two-dimensional space. Tissue is then placed on top of the surface, and the resident mRNA is brought into contact with the primer, either by diffusion of RNA from the tissue to the surface or diffusion of the barcoded primers into the tissue. Typically, primers with poly(T) sequences on their 3′ termini are used to capture mRNAs across the transcriptome.

Spatially indexed pixels have been generated in a variety of ways. The first strategy, used by Stahl and colleagues[17] and used in the Visium technology sold by 10x Genomics, uses a microarrayer spotting robot to deliver a unique barcode to a fixed, known location on the surface of a slide. These spots are 50–200 µm in size (although upcoming Visium products will reportedly have smaller pixels) and are separated by a similar amount of white space to prevent mixing during liquid handling. A second strategy uses solid microparticles for spatial barcoding. In Slide-seq[13], beads that are 10 µm in diameter are used as the solid support for oligonucleotide synthesis, in which the bead barcode is created by split-pool cycles, a process first developed for single-cell barcoding[22]. The beads are fabricated into a tightly packed uniform monolayer on a slide, and the locations of each barcode are ascertained by in situ sequencing. Alternative strategies, either for the method of affixing barcodes to beads or for DNA synthesis, enable the pixel size to be reduced to the 1- to 5-µm range[24]. A third strategy is to locally amplify unique barcode sequences by rolling circle amplification (RCA) or bridge amplification. In Stereo-seq, DNA nanoballs are generated by RCA that span ~200 nm in diameter with 500- to 715-nm center-to-center spacing, and each colony barcode is sequenced. A poly(T) capture sequence is then ligated onto the barcoded nanoballs to enable capture of released RNA. In Seq-Scope[25], barcoding is accomplished by local bridge amplification of DNA randomers directly onto an Illumina sequencing flow cell to create colonies that are 0.5–1 µm in diameter. Pixels can also be formed combinatorically. In DBiT-seq[26], microfluidic channels deliver barcoded reverse transcription primers to the RNA within tissue; the channel apparatus is then rotated 90° to deliver a second set of primers in the orthogonal direction that are ligated in situ, creating a paired barcoding scheme for recovering two-dimensional coordinates.

Once surface barcoding is complete, the downstream workflows of these technologies are remarkably convergent. In most technologies, tissue is placed in contact with the barcoded surface, and mRNA diffuses to the barcoded primers (with the exception of DBiT-seq, in which primers are diffused into the tissue; Fig. 2b). Reverse transcription, cDNA amplification and short-read sequencing generate reads that contain mRNA fragments for transcriptome alignment paired with barcode sequences that are matched back to the pixel whitelists. The result is the spatial localization of each detected transcript to each pixel.

**Quality control.** In sST, the two most important quality parameters are the mRNA capture sensitivity per unit area and the spatial accuracy of mRNA detection, which can be reduced by simple

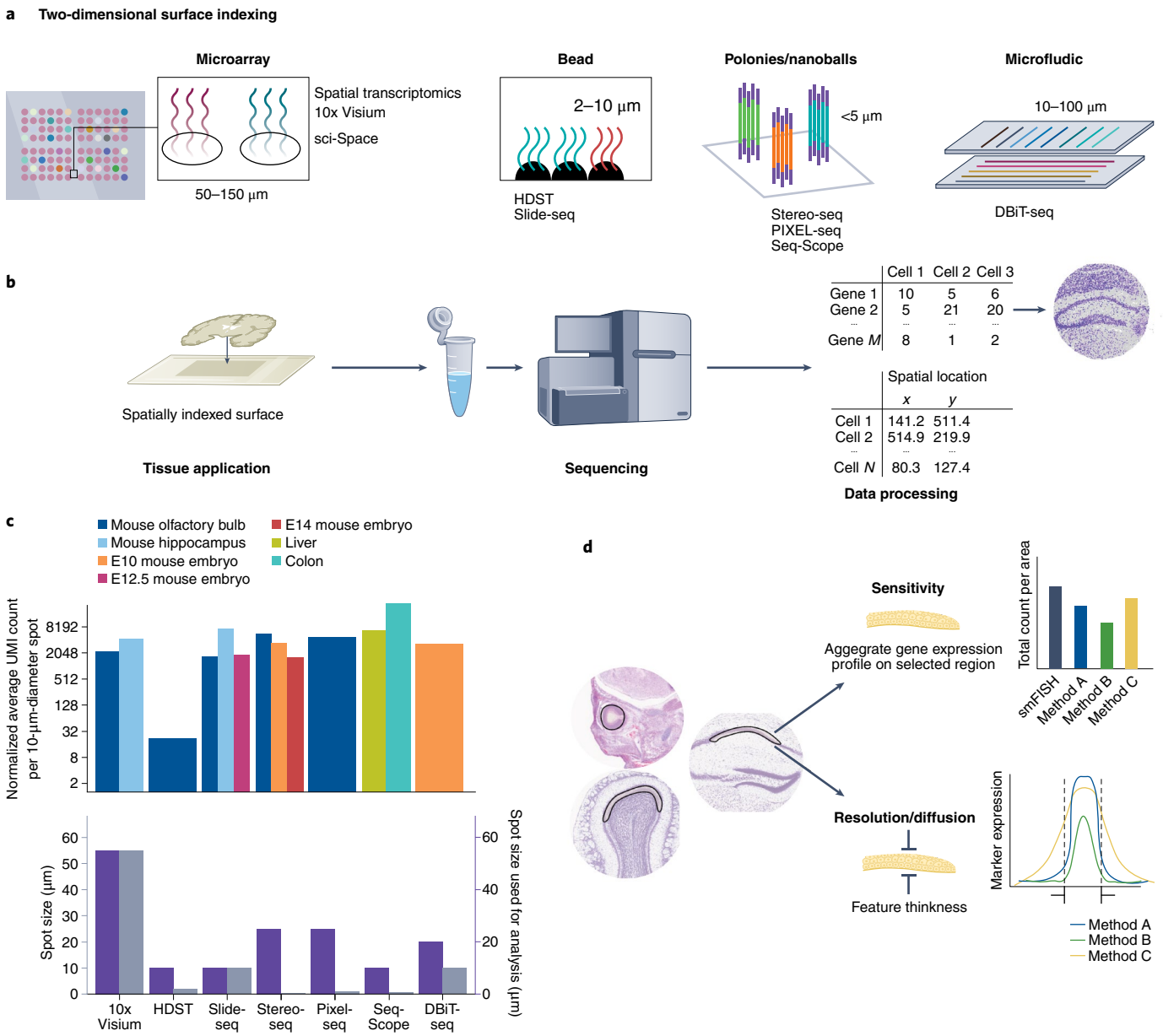


**Fig. 2 | sST methodology and characterization. a**, Workflow of sST methods. Strategies for the fabrication of indexed pixel surfaces where DNA-barcoded primers are associated with spatial localization. Microarray-based strategies use deterministic DNA barcodes printed on glass slides. Bead-based strategies use DNA-conjugated beads with diverse, clonal barcodes whose spatial locations are ascertained. Nanoball- or polony-based strategies use local clonal amplification to generate clusters of clonally barcoded primers. Microfluidic barcoding uses channels to deterministically deliver row and column barcodes to a tissue, forming a two-dimensional grid. HDST, high-definition spatial transcriptomics. **b**, Steps of sequencing library generation downstream of surface indexing. Basic computational processing of the data results in a digital gene expression matrix with a paired table of coordinates for each pixel. **c**, Top, sensitivity of selected sST technologies[16,17,24–27,88] represented by the average number of UMI counts per normalized 10-μm spot and colored by the tissue type. Bottom, the reported spot size for each technology is shown in gray, and the binned resolution used for analysis is shown in purple. **d**, Proposed approaches for benchmarking sST technologies. Representative regions of interest (outlined areas in images on the left) are selected from tissues, and different quality control metrics are applied to the selected region to quantify the sensitivity of RNA capture and the spatial resolution. A marker gene specifically expressed in selected regions was chosen. For capture sensitivity (bar plot, top right), the total counts of the marker gene are summed within the selected region and compared to the counts ascertained by an smFISH reference assay. For resolution (density plot, bottom right), the intensity of marker gene expression is quantified across a dimension of the feature, and the feature thickness is compared to the full-width at half-maximum of the profile. Methods A, B and C represent theoretical ST technologies.

lateral diffusion or more complex technical artifacts. In addition, characterizing and comparing sST technologies involves several additional parameters. One is the sequencing efficiency (or the amount of DNA sequencing required to ascertain a particular number of mRNA molecules). Another is the spatial area covered by the technology (and its flexibility to accommodate tissues of different sizes, compositions and shapes), which can be quite germane to a technology's applicability to specific biological problems.

The capture sensitivity of a technology is highly influenced by the specific molecular and cytoarchitectural features of the tissue

being assayed. Tissues can vary by several orders of magnitude in how much RNA can be captured per unit area because of differences in RNase content, cell density, extracellular matrix composition and other features. To date, the performance of new technologies has not been assayed with a set of consistent, standardized tissues (Fig. 2c shows select sensitivity and resolution metrics across a variety of technologies and tissue types). To properly characterize, benchmark and compare new technologies and improvements, it will be highly beneficial for the field to establish a set of reference tissues with well-defined histological structures that test different technological challenges (Fig. 2d). Stahl et al. validated their technology with the adult mouse olfactory bulb, which has been subsequently used by several other sST technologies for validation. The olfactory bulb is composed of five discrete layers, each with well-known molecular markers, making it a useful model for technology validation. In addition, the use of mouse embryos of different ages is reasonably common[16,26,27], as they have extensive cytoarchitectural variation in small regions, and the cells are smaller and more compact than those in the brain. We propose the additional inclusion of an embryonic day 12 (E12)–E14 mouse embryo sectioned sagittally at a position along the medial–lateral axis that includes the eye. The eye is clearly identifiable, even by non-experts, and is a symmetric structure with clear histological layers, making direct comparisons across experimental replicates and regimens more straightforward. The consistent usage of these two standardized tissue types and other tissues with similar features would be enormously helpful to overall technology development goals in the field.

To quantify mRNA capture, a histological structure can be segmented from an ST-derived pseudoimage, generated either from shading individual pixels by the number of unique molecular identifiers (UMIs) or by plotting the intensity of a metagene that correlates with the chosen histological structure[13] (Fig. 2d). Once segmented, counts of individual genes are summed within the area; single-molecule FISH (smFISH) performed on the same genes in the same histological structure provides a direct, rigorous comparison of a technology's sensitivity. If the histological structure is composed of a uniform cell type, it is also possible to compute an average expression per cell by counting cells in the feature for direct comparison to single-cell sequencing data.

From this same analysis, the spatial resolution of the technology can be assessed. A dimension of the segmented histological structure can be extracted by measuring the distance between full-width at half-maximum of intensity. That distance is then compared to the same dimension measured with an image of the same feature generated by histological or smFISH staining. This provides a quantitative assessment of lateral diffusion of transcripts from their source within the tissue. A similar approach can be taken with the same data to quantify false-positive noise (although this is less commonly a problem in sST technologies because individual genes known to be excluded from specific histological structures are quantified within that excluded feature in the ST dataset and compared to counts of that gene from smFISH data). These simple analyses, if performed on widely available agreed-on tissue samples, provide a common language for characterizing and comparing sST technologies.

## iST

In parallel to the development of sST technologies, there has been an explosion in the number of iST approaches.

In Fig. 3, we outline the broad concepts behind iST. In iST, RNA molecules are specifically tagged with fluorescent probes by complementary hybridization. These probes are then imaged using fluorescence microscopy. Although fluorescence-based RNA detection in situ has been widely used for more than two decades, spectral limitations have prevented routine simultaneous imaging of roughly more than five to ten distinct organic fluorescent molecules. To overcome this limitation, recent iST approaches generally use multiple sequential imaging rounds and combinatorial strategies for detection of transcripts. Thus, a specific iST approach is largely defined by the detection modality, by how RNAs molecules are labeled and the multiplexing approach or by how multiple RNA transcripts are detected across sequential imaging rounds. These approaches have primarily been driven by advances in three fields: oligonucleotide synthesis[28], fluorescence microscopy and single-cell transcriptomics. Recent advances in oligonucleotide synthesis now enable specific synthesis of $10^5$–$10^6$ individual oligonucleotide sequences in a pooled fashion, critical for generation of barcoded hybridization probes. Recent developments in sensitive scientific complementary metal–oxide semiconductor detectors[29,30] and organic fluorophores now enable sensitive, high-throughput detection of labeled RNAs in cells and tissues. Lastly, comprehensive single-cell atlases allow for the selection of informative RNA subsets for labeling.

**Detection modality.** Three main strategies for labeling RNA molecules in situ are used in iST: direct probe-based detection, enzymatically assisted probe-based detection and direct enzymatic sequencing of RNA molecules in situ (Fig. 3a). All detection modalities start with fixed cells and tissues, where the RNA molecules are cross-linked to the cellular matrix, thereby fixing their positions throughout processing.

The first approach, direct probe-based detection, is based on smFISH protocols pioneered by Singer and colleagues[31,32] and subsequently by Raj et al.[31–33] in which RNA transcripts are tiled with many (>20) short (20–50 nucleotides) fluorescently labeled complementary oligonucleotide probes recruited to a single diffraction-limited spot, generating a punctate high-specificity signal.

The second approach, enzymatically assisted probe detection, generates sensitivity and specificity through enzymatic detection and polymerase-based amplification with RCA[21,34]. There are several advantages to enzymatic-assisted probe detection compared to direct probe-based detection methods. First, RCA amplicons are bright and detectable with a high signal-to-noise ratio in fluorescence microscopy, even with lower magnification and exposure times. Second, enzymatic gapfill through reverse transcription allows de novo sequence on the RNA to be targeted and sequenced and enables interrogation of genetic variation (for example, single-nucleotide polymorphisms) and barcodes[21,35,36]. Third, increased signal-to-noise ratio allows more diverse iterative barcoding approaches, such as commercial fluorescence sequencing chemistries. The initial methodology circularized probes hybridized to in situ-generated cDNA, but more recent methods have amplified directly from RNA to increase detection efficiency[4,37].

The third detection approach, direct enzymatic sequencing, uses in situ enzymatic reactions to perform RNA-sequencing library construction within cells and tissues. This approach, pioneered by Church and colleagues[20,37] and subsequently expanded by Boyden and colleagues[20,37], uses in situ reverse transcription with random hexamer primers to generate cDNA, which is subsequently fragmented and intramolecularly circularized. These circularized molecules are subsequently amplified via RCA. Here, specificity for individual RNA molecules is conferred through alignment of in situ sequencing reactions (see 'combinatorial multiplexing') performed on the RCA amplicons. Because of its untargeted nature, this approach is most similar to sST and offers the possibility of hypothesis generation from transcriptome-wide data. However, the low conversion rate of RNA molecules to sequenced RCA amplicons limits its application in many ST experimental contexts that require sensitive gene expression quantification.

**Multiplexing.** There are two main classes of iST-based multiplexing: sequential readout and combinatorial multiplexing (Fig. 3a,b). Both classes leverage multiple imaging rounds to overcome the limitations in spectral bandwidth.

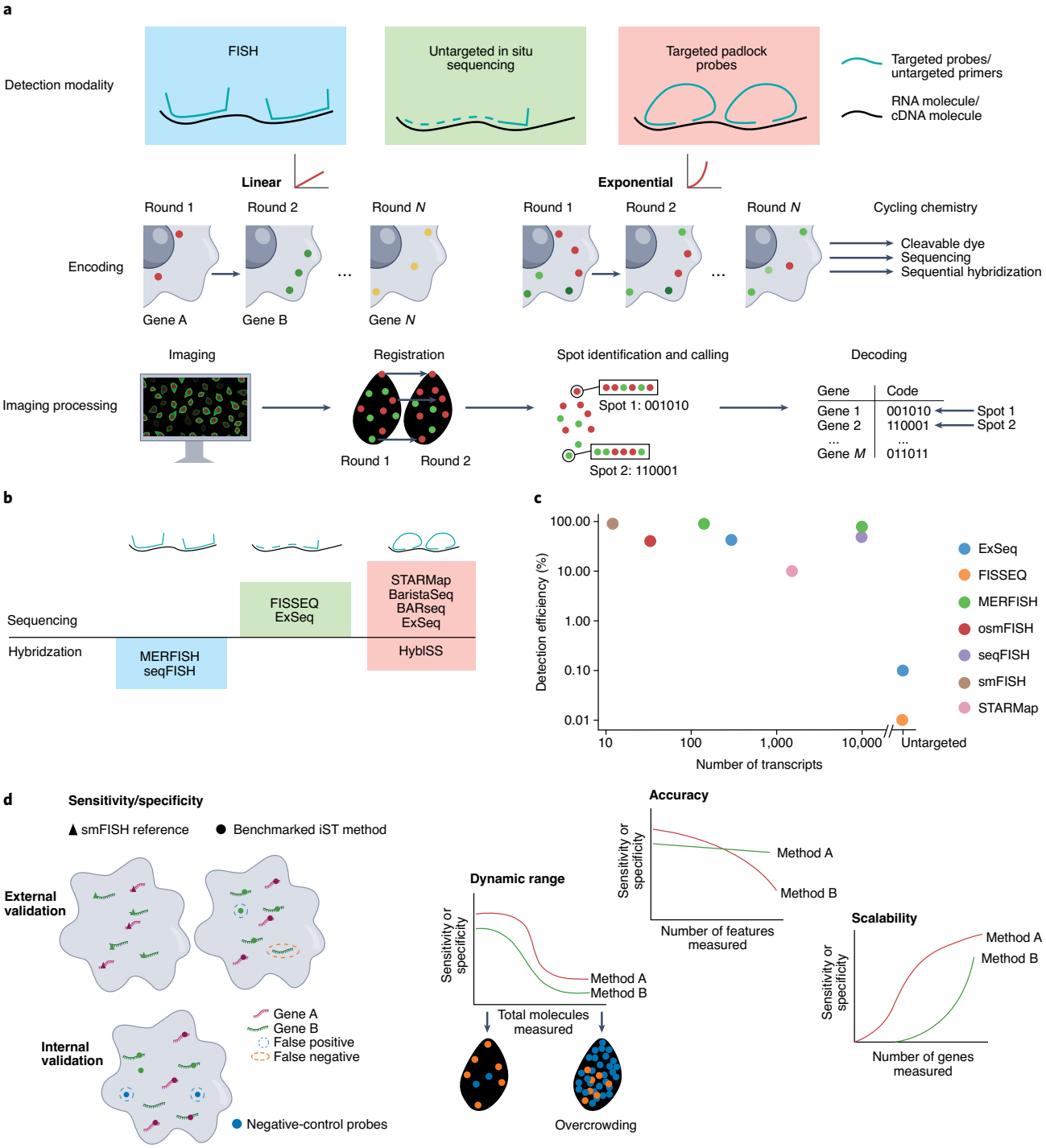


**Fig. 3 | iST methodology and its characterization. a**, Depiction of the three fundamental steps in iST. Targeting chemistry summarizes how the target mRNA is labeled; black lines represent mRNA molecules, and blue lines indicate oligonucleotide probes. Encoding summarizes two strategies for gene encoding to enable multiplexing. Linear encoding labels different mRNAs in each imaging round. Exponential encoding labels each mRNA in multiple imaging rounds. Image processing highlights major steps in downstream image processing after data collection. First, samples are registered between imaging rounds to the same coordinate space. Then, spots corresponding to single RNA molecules are identified and assigned to imaging rounds. Last, gene identity is decoded for each spot based on the imaging rounds. **b**, Summary of different encoding and targeting chemistries and key methods in each category. **c**, Summary of the detection efficiency of select iST technologies at different numbers of genes simultaneously measured[18,34,37,38,40,42,43,89]. **d**, Experiments to measure and compare the performance of methods and corresponding quality control metrics. Sensitivity and specificity experiments assess the rate of false positives and false negatives through comparisons with smFISH (external validation) and internal positive- and negative-control transcripts (internal validation). As the number of molecular features increases in iST, quality control in dynamic range, accuracy and scalability need to be examined. Dynamic range represents the maximum number of molecules that can be measured, accounting for molecular crowding. This may differ between methods as a function of signal to noise and the sparsity of codebooks, and it is dependent on the accuracy of the measurement. Scalability considers how the experimental time and cost scales with the number of features.

In sequential readout approaches, a unique set of mRNA molecules is labeled and imaged, and the fluorescent probes are removed in each imaging round. Subsequently, in the next imaging round, a new set of mRNA molecules is labeled, and the process is repeated. In this way, the number of unique mRNA targets imaged scales as number of rounds × number of fluorescent channels. As one example, in non-barcoded and unamplified cyclic-ouroboros smFISH[38], smFISH probes are stripped by denaturants during each imaging step, followed by a new round of smFISH staining. The main advantage of sequential readout methods is that they are simple to implement because they do not need sophisticated image processing and alignment over many imaging rounds, and they are robust to encoding errors and to higher densities of labeled RNAs. However, this ease of implementation comes with a substantial trade-off in multiplexing scalability up to roughly 100 genes, as sample stability over repeated imaging rounds becomes limiting (Fig. 3c).

In combinatorial multiplexing approaches, each mRNA molecule is interrogated over multiple imaging rounds, and its identity is decoded by the combination and order of the images in which it is (and is not) detected[18,19,39–41]. For example, in multiplexed error robust FISH (MERFISH), Zhuang and colleagues[39] assigned a binary code to each mRNA molecule interrogated, wherein a 1 corresponds to an imaging round where the mRNA molecule is labeled and a 0 corresponds to an imaging round where it is dark. Here, the codebook size in theory scales as $2^N$, where $N$ is the number of imaging rounds. Critically, Zhuang and colleagues[39] implemented a hamming distance-corrected error-robust codebook, which uses bits in the diversity of the codebook to enable detection and correction of errors in single-imaging rounds. This allowed 140 genes to be imaged in 16 rounds, with the ability to correct for errors in 1 round and the ability to detect if an error had happened in 2 imaging rounds. Several implementations exist under the umbrella of combinatorial multiplexing, which mainly differ in cycling chemistry (Fig. 3a, cycling chemistry). The simplest cycling chemistry is reversible hybridization of a fluorescently labeled probe using heat/denaturants and/or DNAse to remove DNA probes after imaging[38,41]. To increase the speed of cycling, reducing agent-cleavable dyes conjugated to detection oligonucleotides have been used[40]. Lastly, cyclic fluorescent sequencing chemistries, such as sequencing by ligation and sequencing by synthesis, have been directly used in situ for combinatorial readout, primarily for RCA amplicons[4,35,37]. Given that the encoding space of combinatorial multiplexing scales exponentially, the main barrier to increasing the number of genes detected is density of molecules labeled. Recent advancements have enabled detection of 1,000–10,000 or more genes with high efficiency through either sparsification of the codebook or expansion microscopy, with a decrease in throughput due to increased imaging rounds (Fig. 3c)[3,4,18,42,43]. We summarize the combinations of read and detection methods used and highlight the degree of multiplexing and detection efficiency of non-exhaustive technology examples in Fig. 3b,c, with more technologies being developed recently[44–48].

**Image processing.** Following the collection of multiround image data, three primary steps are used in image processing to generate primary ST data: spot detection, image registration and decoding into spatial mRNA localizations. During spot detection, local maxima detection is used to localize the centroid of fluorescent spots corresponding to individual mRNA molecules. In both sequential readout and combinatorial multiplexing approaches, the same cells and tissues are imaged over many imaging rounds. As such, a critical aspect of the multiplexing approach is to align each imaging round to the same coordinate framework. This is generally performed using image features, such as fluorescent nuclear staining (for example, with DAPI), fiduciary markers (for example, fluorescent beads) or the mRNA molecule localizations themselves. For combinatorial multiplexing, the alignment needs to be accurate to the resolution of individual mRNA localization (~single diffraction-limited spot) for decoding. Lastly, following image alignment and registration, for combinatorial multiplexing, the order of fluorescent signals for each localized spot is decoded either through matching to a codebook in multiplexed FISH and targeted in situ sequencing approaches or through matching to the transcriptome for untargeted in situ sequencing. For transcriptome mapping, although in situ sequencing is limited in read length, pioneering approaches have paired sequencing of the same molecules with ex situ high-throughput sequencing to enable longer read lengths.

**Quality control.** In iST, just as in sST, the two key data quality parameters are sensitivity and specificity. These parameters need to be assessed given the specific conditions for detection, multiplexing and image processing for a given technology. Sensitivity and specificity can be assessed by direct external validation via smFISH or by the use of internal positive and negative controls. Several studies using MERFISH or expansion sequencing have directly benchmarked multiplexing performance against smFISH for the same images. The near-quantitative detection rate and low false positivity rate of smFISH allows for direct measurements of detection efficiency (false negative) and false positives. In addition to direct external validation, the use of built-in positive controls can provide internal validation for multiplexed imaging experiments. These include built-in codebook controls with no probes assigned to measure decoding accuracy, scrambled probes to measure false-positive detection rates and standard control mRNA targets to allow for comparison of detection efficiency across experiments (Fig. 3d).

An underexplored aspect of imaging-based approaches is how performance varies with the degree of multiplexing. As the number of interrogated mRNAs increases, molecular crowding can prevent efficient detection in the case of enzymatic approaches and make image processing and decoding difficult for FISH-based detection. In MERFISH[39] and in situ sequencing[37], such molecular crowding has been addressed through expansion microscopy[49,50] at the expense of volumetric throughput. In addition, as more genes are encoded, longer codebooks are needed or fewer bits can be devoted to error correction. As such, a key benchmark needed is how sensitivity and specificity scale as a function of number of genes and spatial molecular density (Fig. 3d).

Lastly, there is a need to systematically assess how image-processing pipelines affect data outputs. This is hampered by the fact that current analysis pipelines for iST are boutique and technology specific. In addition, a current lack of standardized data formats across the field for images makes it difficult to develop and benchmark generalized iST image-processing pipelines (from registration, spot calling and decoding to cellular segmentation). As such, generalizable, open-source image-processing tools for iST (https://spacetx-starfish.readthedocs.io/en/latest/) are needed, together with standardized primary datasets on which to benchmark such tools. Ideally, these benchmarks and quality control parameters need to be explored in the context of a set of standardized reference tissues, preferably shared with sST approaches outlined in Fig. 2.

## Beyond ST in cells and tissues

The recent and rapid progress in the development of ST methodologies is a harbinger of a broader technological transformation of genomics to the study of intact cells and tissues. The same tools that enabled ST (high-throughput DNA sequencing, novel barcoding strategies, new microscopy tools and innovations in molecular enzymatics) will increasingly allow genomics to be used to answer questions in cell and tissue biology. We anticipate technology development proceeding in three main domains (Fig. 4).

**Applications to other modalities.** Modern genomics has assembled an impressive array of measurement technologies of biomolecules

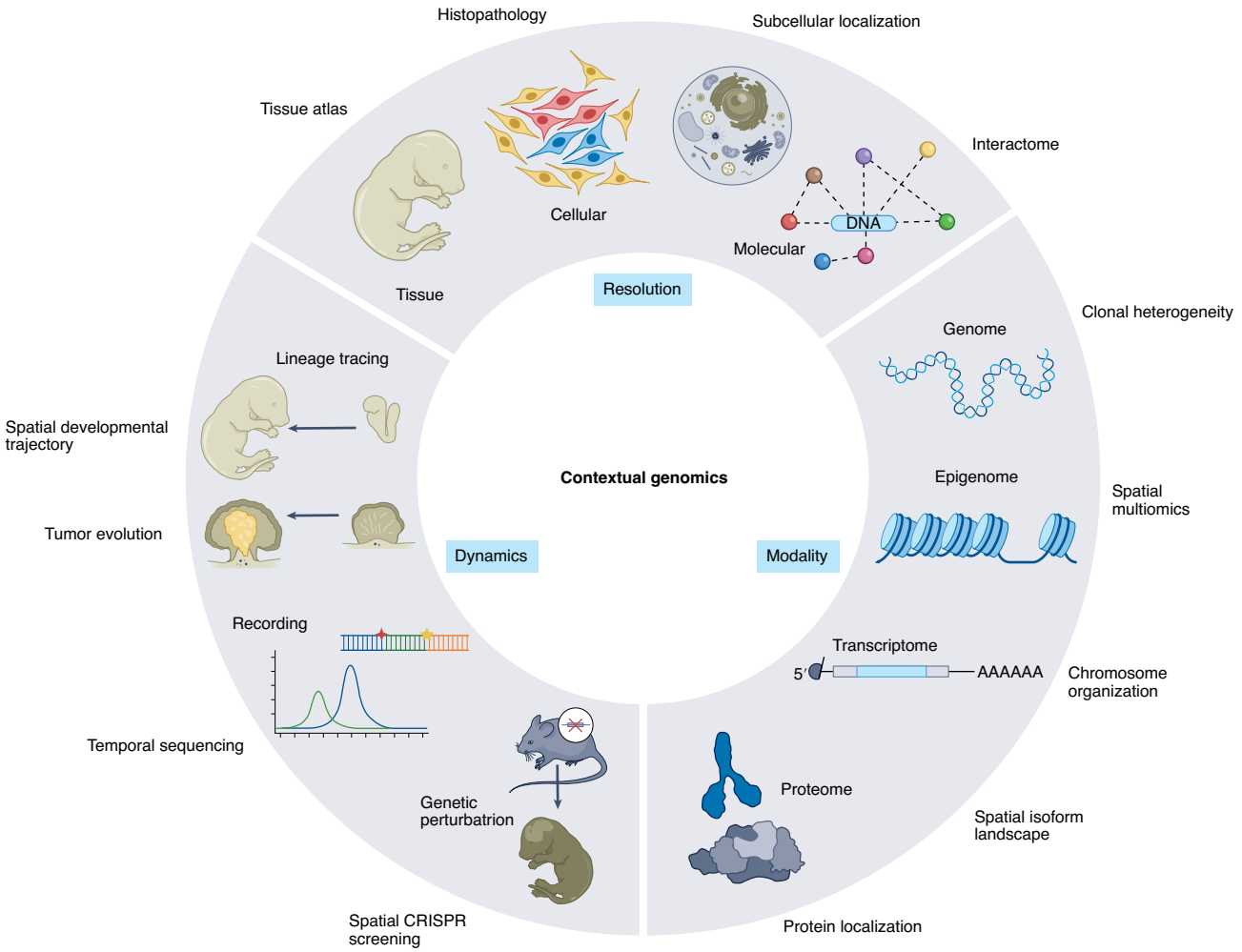


**Fig. 4 | The future of contextual genomics.** The in situ measurement of transcriptomes marks the beginning of a larger technological effort to import genomic measurements into intact biological systems. Innovation along three major axes (resolution, modality and dynamics) will provide powerful new tools for interrogating tissue structure and function.

through creative combinations of enzymatic and biochemical manipulations with DNA-sequencing readouts. However, most of these technologies can only be used in vitro on substrates extracted and purified from tissues or cells. Adapting genomics technologies to function within intact tissue sections represents an enormous opportunity for biological discovery. We anticipate that the progress ST technologies have made in digitally counting transcripts will be increasingly applied toward other modalities of genomic measurements. For example, iST has recently been used to quantify nascent mRNA by targeting introns[51], and this approach could be similarly adapted to quantify splice variants by targeting exon–exon junctions[51]. The discovery of isoform variants currently requires long-read technology and/or plate-based methods to assess. Future adaptation of long-read pipelines or targeted isoform sequencing in situ should enable scalable adaptation to the spatial domain. Although the immense diversity of isoforms across cell types is well known, many questions remain concerning how this variability is functionally manifested across cell types and tissue localizations.

Beyond the transcriptome, spatial profiling of genomic variation has begun to be explored, but tools are still in relative nascency. Available technologies have primarily measured genome structure through imaging[52–55], but more recently, sequencing-based strategies have also been reported[56]. Many of the same techniques used by these technologies could also be applied to study epigenomic modifications and regulation[57]. These techniques would have applications in cancer mutational profiling, and spatial profiling of genomic variation may help elucidate the functional relevance of somatic mutations in aging and disease.

Spatial methods may also offer an opportunity to unify the world of genomics and proteomics. Although genomics has traditionally been focused on dissociated measurements, interrogation of protein localization with affinity reagents has long been performed with low multiplexity in situ. Recent developments in DNA barcoding of affinity reagents, such as antibodies, have enabled highly multiplexed protein readouts via sequencing[26,58]. These approaches are readily adaptable to the spatial domain, especially in the context of spatial-capture ST measurements. In the near future, whole-proteome affinity libraries (including antibodies, nanobodies and aptamers) may be used and interrogated in situ. Furthermore, proximity-based enzymatic reactions, such as ligation[59] and polymerase extension[60–62], may enable high-throughput protein–protein, DNA–protein and RNA–protein interaction measurements within tissues. Lastly, there has been rapid development in numerous approaches for novel protein-sequencing methods that leverage single-molecule imaging[63]; one day, this may allow direct protein sequencing in tissues.

**Improvements in resolution.** Current spatial genomic approaches span a wide range of spatial resolution from broad tissue regions to subcellular localization, with an overall trade-off between volumetric throughput of tissue and spatial resolution of data collected. Different spatial resolutions are suited to different classes of biological problems. The ability to perform cellular segmentation on the measured transcripts is crucial for many downstream applications, such as quantifying cell-type composition and organization of tissues. In both sST and iST, segmentation is a computational problem. For sST, it involves the deconvolution of cell-type mixtures from a pixel[64–66]; the problem is made much simpler when pixel size is reduced to the size of individual cells. For iST, the computational task is to assemble individually detected transcripts into cells from microscopy images[67–69]. Progress on both of these efforts has been made, but additional computational innovation will greatly accelerate the ability of ST tools to be applied to problems in tissue biology[70].

After tissue composition, increasing resolution will enable models of tissue organization that take into account cellular interactions. Fundamental questions include how does one extract the relevant cellular networks that compose tissues? And, how does one discover the functional receptor–ligand interaction networks between cells? Such questions are most suited to technologies with cellular resolution or higher to enable accurate assignment of molecular signatures to networks of interacting cells.

Additionally, technologies that enable the precise in situ colocalization of biomolecules (potentially at resolutions exceeding the diffraction limit) remain almost entirely unexplored. One intriguing exception is the proposal by several groups to use in situ PCR amplification with local concatenation to detect spatial proximity between two nucleic acids[60,61]. A method for in situ molecular colocalization would be enormously biologically enabling, from the quantification of transcription factor binding sites to the colocalization of biomolecules to specific organellar compartments or the quantification of gap junctions or other cell–cell interactions. Although genomic technology for quantifying such proximity or interaction events does exist, such as chromatin immunoprecipitation–sequencing or ribosomal profiling, these methods are most suited to bulk-level analyses of lysed tissue rather than analysis in situ.

**Observation of cellular dynamics.** Almost all current genomic technologies are end-point measurements that provide a snapshot of biology, but there is an exciting future for approaches that provide temporal context to spatial genomic measurements. Looking forward, we anticipate innovations related to integration of molecular recording technologies, computational inference of dynamics and in situ perturbations.

The intersection between spatial genomic technologies and synthetic genetic manipulation approaches for molecular recording is especially promising. Recently, a burgeoning number of gene-editing-based molecular recording technologies that encode lineage and signaling dynamics into genomic sequence have been described[71–76]. These approaches have leveraged single-cell sequencing for subsequent readout of the genomic information. Given the inherent similarities in chemistry between single-cell and sST, genomic recordings of lineage may be easily translated to the tissue context. Such measurements will be important for developmental biology, wherein information regarding cellular lineage, cell state and adult tissue organization can be simultaneously combined to form the developmental picture. In tumor evolution, such approaches can answer important questions regarding the spatial heterogeneity of tumor clones and the relationship between tumor clone fitness[56,77] and the cellular microenvironment. Beyond lineage, novel recording technologies are beginning to encode cellular histories and transcriptomic states[73], which may offer the promise of fully four-dimensional measurements of cell states with genomics.

Regarding the computational inference of dynamics, there has recently been an explosion in the number of computational tools for single-cell studies, such as pseudotime[78,79] and RNA velocity[80,81]. Approaches developed for single cells have been successfully applied to ST data, such as the application of RNA velocity to ST of the developing mouse cortex[82]. However, future development of spatial-oriented toolboxes for computational dynamics will more directly leverage the unique aspects of ST data. One may leverage the subcellular localization of RNA to infer additional temporal information regarding dynamics (for example, RNA is made in the nucleus but translated in the cytoplasm[42]). More importantly, spatial information may provide contextual ground truth in pseudotime analyses, such as using cellular localization to assign nodes in cellular trajectories.

Lastly, perturbations allow causality to be inferred from end-point measurements. Recent approaches have enabled nucleic acid barcoding of both genetic and chemical perturbations[83] for single-cell readout. Similar to molecular recording, perturbation barcoding approaches[83–85] may be inherently compatible with spatial-capture methods. Additionally, barcode readout via in situ sequencing has been demonstrated for imaging methods[36,86,87]. Analyzing the effects of genetic perturbation in situ using ST will enable profiling of phenotypes that cannot be accessed in the absence of tissue context. These phenotypes are numerous, including cellular localization and cell–cell interactions.

## Conclusions

The application of genomics to tissues represents an exciting, multifaceted domain of technology development. Faced with many exciting opportunities for new measurements in the future, what can the field do to further accelerate the pace and quality of these developments? The early and easy sharing of data and protocols are key catalysts of technological progress. This lesson is perhaps best exemplified by the recent rapid progress in the field of single-cell genomics, where manuscripts are usually preprinted and processed, raw data are shared with publication (and often before) and protocols are made easily accessible through online portals. In ST, the heterogeneity of methodologies breeds a large variety of file formats and data structures. This makes data and protocol sharing more challenging but also all the more necessary to contextualize a new method within the broader technological ecosphere. Efforts have begun to develop systematized pipelines for image data processing (https://github.com/spacetx/starfish), but more needs to be done to build generalizable file formats and standardized data storage and access pipelines. Relatedly, we see the adoption of standardized tissues and quality control measurements as essential to developing a common language for technological progress in ST. Without such standards, a clear comparison of even basic metrics, such as the capture efficiency of RNA, is not possible.

The advent of spatial genomics also poses an additional important challenge of how to facilitate hypothesis generation and testing by biological experts using these powerful new tools. Undoubtedly, a key to this will be invention and adoption of computational tools for analysis of these new data types. Computational tools will be important for not only extracting biological insights from the data but also informing the design of spatial genomics experiments. We anticipate that the maturation and dissemination of spatial genomics technologies will become a critical driver of biological discovery across many fields of tissue and disease biology.

Received: 2 February 2022; Accepted: 26 July 2022;
Published online: 03 October 2022

## References

1. Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).

2. Ortiz, C. et al. Molecular atlas of the adult mouse brain. *Sci. Adv.* **6**, eabb3446 (2020).
3. Zhang, M. et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).
4. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
5. Chen, R. et al. Decoding molecular and cellular heterogeneity of mouse nucleus accumbens. *Nat. Neurosci.* **24**, 1757–1771 (2021).
6. Knoten, A., Urata, S., Naik, A. S., Eddy, S. & Zhang, B. An atlas of healthy and injured cell states and niches in the human kidney. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.28.454201 (2021).
7. Ferreira, R. M. et al. Integration of spatial and single cell transcriptomics localizes epithelial–immune cross-talk in kidney injury. *JCI Insight* **6**, e147703 (2021).
8. Marshall, J. L., Noel, T., Wang, Q. S. & Bazua-Valenti, S. High resolution Slide-seqV2 spatial transcriptomics enables discovery of disease-specific cell neighborhoods and pathways. *iScience* **25**, 104097 (2021).
9. Asp, M. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660 (2019).
10. Chen, H. et al. Dissecting mammalian spermatogenesis using spatial transcriptomics. *Cell Rep.* **37**, 109915 (2021).
11. Madissoon, E., Oliver, A. J. & Kleshchevnikov, V. A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.26.470108 (2021).
12. Chen, W.-T. et al. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* **182**, 976–991 (2020).
13. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
14. Wu, Y. et al. Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discov.* **12**, 134–153 (2021).
15. Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I. & White, R. M. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat. Commun.* **12**, 6278 (2021).
16. Stickels, R. R. et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
17. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
18. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
19. 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).
20. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
21. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
22. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
23. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
24. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
25. Cho, C.-S. et al. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184**, 3559–3572 (2021).
26. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681 (2020).
27. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).
28. Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
29. Huang, Z.-L. et al. Localization-based super-resolution microscopy with an sCMOS camera. *Opt. Express* **19**, 19156–19168 (2011).
30. Saurabh, S., Maji, S. & Bruchez, M. P. Evaluation of sCMOS cameras for detection and localization of single Cy5 molecules. *Opt. Express* **20**, 7338–7349 (2012).
31. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
32. Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. Single-cell gene expression profiling. *Science* **297**, 836–840 (2002).
33. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
34. Lee, J. H. et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
35. 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).
36. Feldman, D. et al. Optical pooled screens in human cells. *Cell* **179**, 787–799 (2019).
37. Alon, S. et al. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, eaax2656 (2021).
38. Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
39. 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).
40. 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).
41. 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).
42. 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).
43. Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
44. Goh, J. J. L. et al. Highly specific multiplexed RNA imaging in tissues with split-FISH. *Nat. Methods* **17**, 689–693 (2020).
45. Nagendran, M., Riordan, D. P., Harbury, P. B. & Desai, T. J. Automated cell-type classification in intact tissues by single-cell molecular profiling. *eLife* **7**, e30510 (2018).
46. Liu, S. et al. Barcoded oligonucleotides ligated on RNA amplified for multiplexed and parallel in situ analyses. *Nucleic Acids Res.* **49**, e58 (2021).
47. Dar, D., Dar, N., Cai, L. & Newman, D. K. Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science* **373**, eabi4882 (2021).
48. Sountoulidis, A. et al. SCRINSHOT enables spatial mapping of cell states in tissue sections with single-cell resolution. *PLoS Biol.* **18**, e3000675 (2020).
49. Chen, F. et al. Nanoscale imaging of RNA with expansion microscopy. *Nat. Methods* **13**, 679–684 (2016).
50. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015).
51. Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376 (2018).
52. 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 (2020).
53. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
54. Takei, Y. et al. Single-cell nuclear architecture across cell types in the mouse brain. *Science* **374**, 586–594 (2021).
55. Payne, A. C. et al. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* **371**, eaay3446 (2021).
56. Zhao, T. et al. Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature* **601**, 85–91 (2022).
57. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
58. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
59. Söderberg, O. et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000 (2006).
60. Weinstein, J. A., Regev, A. & Zhang, F. DNA microscopy: optics-free spatio-genetic imaging by a stand-alone chemical reaction. *Cell* **178**, 229–241 (2019).
61. Hoffecker, I. T., Yang, Y., Bernardinelli, G., Orponen, P. & Högberg, B. A computational framework for DNA sequencing microscopy. *Proc. Natl Acad. Sci. USA* **116**, 19282–19287 (2019).
62. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res.* **39**, e102 (2011).
63. Swaminathan, J. et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* **36**, 1076–1082 (2018).
64. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
65. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2021).
66. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).

67. Petukhov, V. et al. Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* **40**, 345–354 (2021).
68. Littman, R. et al. Joint cell segmentation and cell type annotation for spatial transcriptomics. *Mol. Syst. Biol.* **17**, e10108 (2021).
69. Prabhakaran, S. Sparcle: assigning transcripts to cells in multiplexed images. *Bioinform. Adv.* **2**, vbac048 (2022).
70. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nat. Biotechnol.* **40**, 308–318 (2022).
71. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
72. Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR–Cas in human cells. *Science* **353**, aag0511 (2016).
73. Rodriques, S. G. et al. RNA timestamps identify the age of single molecules in RNA sequencing. *Nat. Biotechnol.* **39**, 320–325 (2021).
74. Kalhor, R. et al. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).
75. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
76. Chow, K.-H. K. et al. Imaging cell lineage with a synthetic digital recording system. *Science* **372**, eabb3099 (2021).
77. Fennell, K. A. et al. Non-genetic determinants of malignant clonal fitness at single-cell resolution. *Nature* **601**, 125–131 (2022).
78. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
79. Bendall, S. C. et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714–725 (2014).
80. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
81. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
82. Abdelaal, T., Lelieveldt, B. P. F., Reinders, M. J. T. & Mahfouz, A. SIRV: spatial inference of RNA velocity at the single-cell resolution. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.26.453774 (2021).
83. Srivatsan, S. R. et al. Massively multiplex chemical transcriptomics at single-cell resolution. *Science* **367**, 45–51 (2020).
84. Dixit, A. et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
85. Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
86. Chen, X. et al. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179**, 772–786 (2019).
87. Wang, C., Lu, T., Emanuel, G., Babcock, H. P. & Zhuang, X. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc. Natl Acad. Sci. USA* **116**, 10842–10851 (2019).
88. Fu, X. et al. Continuous polony gels for tissue mapping with high resolution and RNA capture efficiency. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.17.435795 (2021).
89. 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).


## Acknowledgements

We thank members of the Chen and Macosko laboratories for helpful discussions. This work was supported by NIH grants R01HG010647, UH3CA246632 and R33 CA246455 (to F.C. and E.Z.M.).

## Competing interests

F.C. and E.Z.M. are consultants for Curio Bioscience, Inc.

## Additional information

**Correspondence** should be addressed to Fei Chen or Evan Z. Macosko.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.