



Quantitative approaches for investigating the spatial context of gene expression

Je H. Lee *

The spatial information associated with gene expression is important for elucidating the context-dependent transcriptional regulation during development. Recently, high-resolution sampling approaches, such as RNA tomography or single-cell RNA-seq combined with fluorescence *in situ* hybridization (FISH), have provided indirect ways to view global gene expression patterns in three dimensions. Now *in situ* sequencing technologies, such as fluorescent *in situ* sequencing (FISSEQ), are attempting to visualize the genetic signature directly in microscope images. This article will examine the basic principle of modern *in situ* and single-cell genetic methods, hurdles in quantifying intrinsic and extrinsic forces that influence cell decision-making, and technological requirements for making a visual map of gene regulation, form, and function. Successfully addressing these challenges will be essential for investigating the functional evolution of regulatory sequences during growth, development, and cancer progression. © 2016 The Authors. *WIREs Systems Biology and Medicine* published by Wiley Periodicals, Inc.

How to cite this article:

WIREs Syst Biol Med 2017, 9:e1369. doi: 10.1002/wsbm.1369

INTRODUCTION

Comprehensive cataloguing of the molecular and cellular heterogeneity is important for classifying regulatory pathways into functional categories; however, they alone cannot reveal how sequence-to-function relationships evolve over space and time. Because cells interpret functional sequences in a cell state-, history-, or environment-dependent manner, retracing their decision-making algorithm requires examining cell identities, lineage, and external cellular interactions across a range of landscapes, respectively (Figure 1). Here, modern high-throughput approaches are becoming indispensable for understanding how cells recapitulate complex form and function using a defined set of conserved signaling pathways, while overcoming or responding to biological noise and perturbations.

*Correspondence to: jlee@cshl.edu

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

Conflict of interest: The author has declared no conflicts of interest for this article.

For the past 100 years, optical microscopy and general tissue stains were the main tools for surveying the tissue landscape; however, elucidating the genetic mechanism required extracting the nucleic acids from pulverized tissues and destroying its spatial information. Recently, several new technologies have emerged to address the tissue heterogeneity, including single-cell sequencing and multiplexed *in situ* probe hybridization.^{1,2} Single-cell RNA-seq (scRNA-seq) can measure the transcriptional heterogeneity and reconstruct the cellular location using the known gene expression pattern in space,^{3–5} while single-molecule fluorescence *in situ* hybridization (smFISH) can directly visualize a modest number of genes simultaneously *in situ*.^{6–9} Other approaches combine geographically defined single-cell isolation,^{10,11} laser-assisted sampling,¹² or serial tissue sectioning¹³ with genome¹⁴ or transcriptome profiling.^{15,16}

In situ sequencing is a set of proof-of-concept technologies that can sequence the DNA or the RNA within fixed cells without destroying their spatial context,^{17–19} and it can be used to directly link spa-

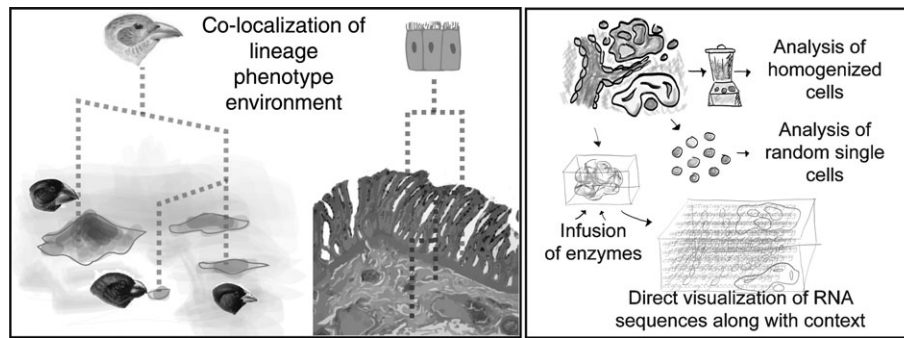


FIGURE 1 | Molecular or cellular taxonomy alone is insufficient for understanding the functional dynamics of genetic and phenotypic evolution, as it requires analyzing how the selection pressure from the environment changes the phenotype from a common ancestor. To properly address this question, one needs to compare multiple cell lineages, cell states/types, and microenvironments in parallel. Traditionally, the genetic material was isolated from pulverized tissues, masking the cellular heterogeneity as well as their spatial context. Methods now exist to sample randomly chosen cells or from spatially defined regions; however, they all lack the precise spatial resolution for understanding cell–cell or cell–environment interactions.

tial features to particular genetic elements in native tissue specimens, as long as sequencing libraries can be constructed from the sequence-of-interest inside the cell. Technologically, *in situ* sequencing is an extension of high-throughput DNA sequencing, in which the nucleic acid sequence is read-out on a solid surface using optical microscopy.^{20,21} In next-generation sequencing (NGS), the temporal alignment of colors from fluorescent amplicons determines the sequence of nucleotide bases, and *in situ* enzyme reactions on the glass surface allow for saturation kinetics and robust molecular imaging, leading to longer reads and higher read densities.

Because FISH also involves fixing the nucleic acids onto a solid matrix, the possibility of NGS inside fixed cells had existed. In 2004, Nilsson and coworkers demonstrated a padlock probe and rolling circle amplification (RCA)-based approach to capture, amplify, and image the DNA with single-nucleotide resolution in fixed cells and tissues.²² Later, the Church laboratory incorporated padlock probes for targeted NGS, suggesting that the Nilsson method could be scaled up for NGS inside the cell.^{23–27} Eventually, the Nilsson laboratory demonstrated targeted RNA detection using *in situ* sequencing to identify a small number of transcripts and mutations in tissue sections,^{17,18} which was later followed by the report of transcriptome-wide fluorescent *in situ* sequencing (FISSEQ) from the Church laboratory.^{19,28}

At the present moment, *in situ* sequencing is at the proof-of-concept stage (e.g. known cancer mutations in tissues,¹⁸ cell culture models,¹⁹ and short barcodes²⁹), unlike smFISH or single-cell sequencing, but it may be able to address questions that other methods cannot if a number of technical limitations

can be overcome. With that in mind, this article will focus on the biological and technological hurdles that are addressed by the previous generation of *in situ* hybridization and single-cell technologies and the challenges that remain. To summarize the history of DNA sequencing and *in situ* methods in a couple of paragraphs will undoubtedly neglect seminal contributions from multiple groups, so this article will focus on technologies that immediately preceded or followed the development of *in situ* sequencing. Beyond providing a historical account, the article will also address several key questions in biology, explore whether current methods are ready to tackle these problems, and conclude with a set of technological goals for investigating the spatial dynamics and the functional evolution of gene regulation during the emergence of form and function in development.

A BRIEF WALK THROUGH MODERN *IN SITU* AND SINGLE-CELL METHODS

The padlock probe-based method from the Nilsson was a conceptually elegant approach to quantifying multiple DNA or RNA molecules with single-nucleotide resolution *in situ*;^{17,18,22} however, the variable sensitivity of individual probes across different samples and gene targets required careful validation and limited its scalability. Moreover, it faced a competition from smFISH, which was more uniformly sensitive and easier to implement, soon making it the gold standard in single-cell analysis *in situ*. With the lower cost of generic oligonucleotide synthesis, it was also more scalable, although it lacked the exquisite single-base specificity of the Nilsson approach^{2,30} (Figure 2).

For decades, FISH was used to localize gene transcripts *in situ*; however, quantifying gene expression was difficult due to the variable signal-to-noise ratio and the nonspecific signal from nucleic acid hybridization. Singer and coworkers demonstrated that smFISH can obtain quantitative measurements of the transcript abundance in single cells and that genetic barcoding can be used to multiplex transcript detection.^{6,32} Raj, along with colleagues in Tyagi and van Oudenaarden laboratories, made smFISH practical and robust by using multiple short probes that co-localize on the transcript for a high signal-to-noise ratio^{7,33,34} (Figure 2). This also permitted better tissue penetration, reducing the need to optimize

fixatives and partial tissue digestion in FISH experiments. In addition, these advances affirmed an important concept in NGS and *in situ* studies: the high specificity and sensitivity can be achieved by co-localizing numerous short reads or probes with quality scores based on a statistical model.

Because counting individual molecules using microscopy is limited by the optical resolution, conceptual approaches in super-resolution microscopy (SRM) were important in inspiring new iterations of smFISH. For example, localization microscopy, such as Stochastic Optical Reconstruction Microscopy (STORM), achieves higher resolution by using stochastic molecular fluorescence over many cycles,^{35–37} suggesting that cyclical imaging of barcoded probes may enable multiplexing in lieu of SRM³⁸ (Figure 3). In fact, Cai and coworkers reported generating a unique temporal barcode by sequentially hybridizing different colored probes to the same transcript, enabling them to multiplex smFISH.⁸ Taking a step further, Zhuang and coworkers used redundancy coding to detect more than 1000 genes using smFISH in an error-resistant manner.⁹ They also demonstrated faster sequential color read-out by inactivating fluorophores with light, thereby shortening the read-out cycle time.

In theory, smFISH can be combined with SRM to profile the messenger RNA (mRNA; ~100,000 molecules per cell);⁴⁰ however, the imaging time required even for a small tissue makes it impractical. Because most biologically relevant gene expression patterns occur across many microns and millimeters, SRM is akin to asking a single Google Street driver to map the whole continent, and it is poorly suited for whole-tissue mapping. One solution could be to have an army of automated drivers, but it would require an enormous amount of time and resources to update the map frequently across multiple continents. Compounding the problem is the fact that many RNA molecules are tightly packaged into subcellular granules, making their quantification using optical imaging challenging at any fixed resolution (Figure 4).

In contrast, a global approach (satellite view) to RNA imaging enables comprehensive and efficient mapping of whole tissues and organs, albeit at much lower resolution. For example, Deisseroth and coworkers developed CLARITY to make whole organs optically transparent for high-resolution three-dimensional (3D) imaging.⁴¹ Since then, other groups have shown faster and easier methods for whole-tissue clearing,⁴² and recently the Deisseroth laboratory and others have demonstrated sequential FISH in optically cleared tissues using serial hybridization of modified RNA targets.^{43,44} Here, a high-

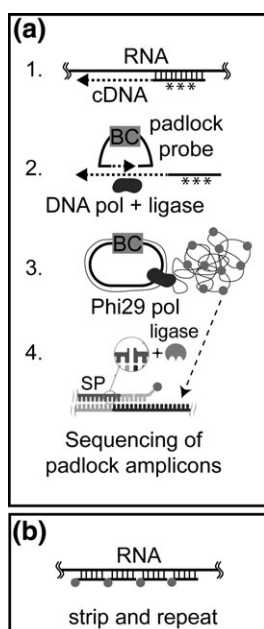


FIGURE 2 | Quantitative methods for detecting multiple RNA molecules *in situ*. (a) The Nilsson method uses target-specific reverse transcription (RT) primers (typically locked nucleic acid (LNA) derivatives) to make cDNA molecules *in situ* used for padlock probe-based T4 DNA ligation.¹⁷ The intramolecular ligation reaction here is less efficient and specific than the sequencing-by-ligation reaction kinetics. The circular padlock probe is then amplified using rolling circle amplification (RCA) that increases the number of barcode-binding sites by 100-fold or more for robust imaging; however, it is not known how the physical constraints or molecular crowding around individual transcripts in tissues affect the RCA bias that is observed in fluorescent *in situ* sequencing (FISSEQ).¹⁹ (b) Compared to the previous method that involves long customized probes and multiple enzymatic steps, single-molecule fluorescence *in situ* hybridization (smFISH) offers the unmatched sensitivity, spatial resolution, ease of use, affordability, and scalability, as long as one can optically resolve individual signals under a microscope.⁷ Because smFISH is so sensitive, this can be challenging for most abundantly expressed transcripts, especially at low magnification.³¹ In addition, smFISH does not have the single-nucleotide specificity of the Nilsson method.

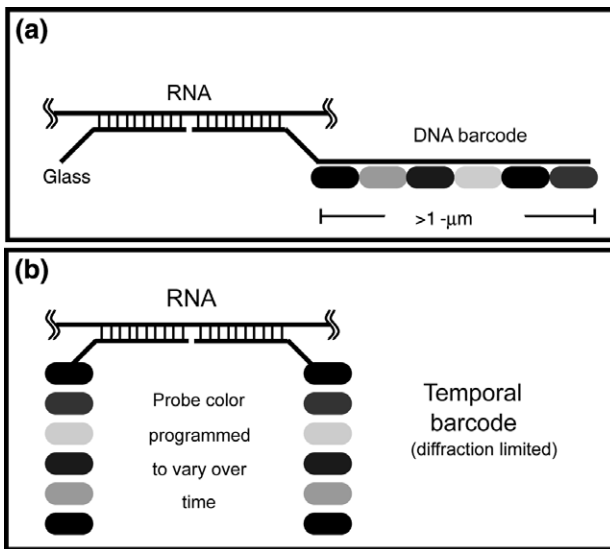


FIGURE 3 | Spatial versus temporal genetic barcoding for multiplexed RNA detection. (a) Genetic barcoding to multiplex transcript detection was first demonstrated *in situ* by Singer and coworkers.⁶ Hood and coworkers then popularized this concept and made it commercially successful for single-molecule RNA quantification (Nanostring, Seattle, WA).³⁹ Here, the target RNA serves as a splint that pulls down a complementary probe with a spatial barcode composed of fluorescent nucleic acid segments (~1–2 μm). Optically resolving various color sequences-associated each RNA molecule enables target identification and quantification. (b) Large barcodes cannot be used for multiplexing in single cells; however, Cai and coworkers showed that targeting the same loci repeatedly with single-molecule fluorescence *in situ* hybridization (smFISH) but using different colors generates a temporal barcode.⁸ Using four-color imaging and seven hybridization cycles, one could interrogate over 16,000 genes in theory, despite a number of practical challenges due to the diffraction limit of optical microscopy and the imaging time required for super-resolution microscopy.

resolution objective for single-molecule imaging is not a practical option for whole-tissue imaging; therefore, RNA detection methods that depend on

single-molecule imaging for multiplexing cannot be used,⁴⁴ although potential solutions are starting to emerge.^{31,45,46}

While smFISH is relatively straightforward to implement, it requires a list of candidate genes, typically based on existing annotations. As NGS has shown, unbiased sequencing often allows for new discoveries compared with targeted gene expression arrays, which are best suited for profiling already known genetic elements. For example, NGS has shown that a large fraction of the long noncoding RNA (lncRNA) is more tissue specific than the mRNA.⁴⁷ In addition, many regulatory enhancers express noncoding RNAs, whose spatial distribution may provide clues for understanding gene regulation.⁴⁸ Because many mRNAs still lack detailed functional annotations, targeted assays are generally limited to well-annotated genes. In addition, massively parallel functional analysis uses expression constructs containing synthetic barcode sequences for *de novo* short-read sequencing.^{49–51}

For these and other reasons, direct RNA sequencing is preferable over targeted hybridization-based methods in many applications (Figure 5). Over the past several years, improvements in the detection sensitivity and reliability have allowed transcriptome-wide scRNA-seq.^{15,16} To investigate the cell-type composition in space, Quake and coworkers used a microfluidics platform for scRNA-seq (Fluidigm, San Francisco, CA), sequencing a couple of hundred single cells from the lung alveoli to infer their location using known anatomic biomarkers.³ Several laboratories took it a step further by integrating high-throughput scRNA-seq with an online database of FISH experiments to reconstruct the cellular composition of 3D tissue structures.^{4,5} These approaches, however, depend on having a database of positional biomarkers, which may not be available in most cases.

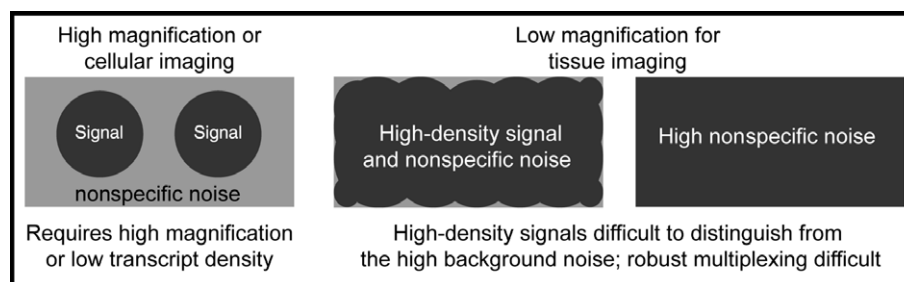


FIGURE 4 | Single-molecule fluorescence *in situ* hybridization (smFISH) requires spatially resolving individual molecules for quantification and multiplexing.^{8,9} The addition of multiple small hybridization probes to cells and tissues generates significant nonspecific fluorescence;⁷ however, the co-localization of many independent probe sequences on the same transcript can be detected as a diffraction-limited spot, resulting in a high signal-to-noise ratio (SNR). But when the transcript density is too high or the imaging magnification is too low, it becomes difficult to discriminate signal from noise, rendering smFISH largely qualitative and challenging for a high degree of multiplexing.³¹

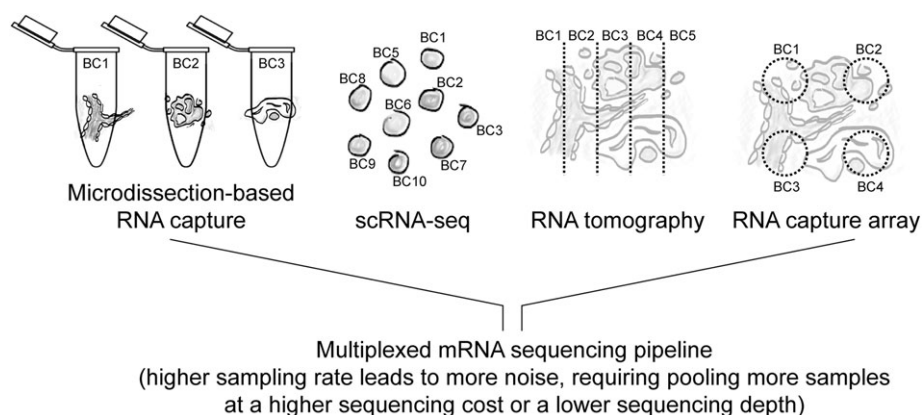


FIGURE 5 | Location-aware sampling methods for next-generation sequencing (NGS). For unbiased profiling of transcriptome-wide gene expression, NGS is currently the only wide available method. Tissue samples can be dissociated into random or sorted single cells and virtually reconstructed later using the known spatial patterns of gene expression (single-cell RNA-seq, scRNA-seq).^{4,5} Alternatively, they can be spatially dissected [i.e., laser capture microdissection (LCM), transcriptome *in vivo* analysis (TIVA)],¹² sectioned (RNA tomography),¹³ or systematically subsampled (i.e., RNA capture array)⁵² for RNA-seq. Here, the practical sampling number limits the spatial resolution as the specimen size increases along multiple dimensions. Practically, only a small fraction of all possible sampling points are used for multiplexed RNA-seq with a lower sequencing depth. In addition, the sampling noise requires pooling multiple regions or single cells together to detect subtle variations, further handicapping NGS-based methods from achieving the single-cell spatial resolution based on less abundant and possibly more tissue-specific transcripts.

To enable *de novo* spatial reconstruction of gene expression, van Oudenaarden and coworkers developed RNA tomography, in which vertebrate embryos or organs were cut into topographically mapped 10–100- μm sections, sequenced, and reconstructed leveraging the scRNA-seq pipeline.¹³ Although one can generate thinner tissue sections, smaller tissue sections lead to more sampling noise, requiring averaging of data from adjacent sections and limiting the spatial resolution to several hundred microns for less abundant transcripts^{1,53} (Figure 5). Recently, a start-up from Joakim Lundeberg's laboratory reported a method of extracting RNA from tissue sections placed directly on a barcoded poly-dT array, followed by NGS and computational reconstruction.⁵² Each spatial element captures $\sim 10\text{K}$ unique mRNA transcripts per 100- μm feature, which is several times lower than that of scRNA-seq. For coarse multiregional sampling, however, this method may provide a less destructive alternative to microdissection (Figure 5).

For higher spatial resolution, Eberwine and coworkers developed transcriptome *in vivo* analysis (TIVA), in which photo-caged biotinylated poly-dT primers are released exactly at the point of laser excitation inside living cells.¹² By pulling down the captured mRNA for sequencing, one could theoretically sequence almost any region in the cell with submicron resolution as long as enough of the identical regions are sampled. If automated, TIVA could be a promising option for studying the transcriptome

composition in different subcellular compartments and cellular regions using unbiased deep sequencing; however, the cell type- and tissue-specific penetration of the TIVA reagent currently limits its application.

To conclude, the imaging resolution, the transcript density, and the imaging time limit multiplexed methods that depend on single-molecule color coding to simpler systems suited for high-resolution microscopy. Multiplexed error-robust fluorescence *in situ* hybridization (MERFISH)⁹ can cycle through multiple probe sets and deduce the transcript identity more accurately than other smFISH methods. Because of its reliance on single-molecule imaging, however, profiling a large number of transcripts at the 'omics' scale is currently out of reach for MERFISH or other smFISH techniques.^{8,9}

Unbiased deep transcriptome profiling is a major benefit of scRNA-seq, but the limited detection sensitivity as well as the sampling noise requires pooling of data from multiple sources to detect small variations. This in turn reduces the sequencing depth for each cell sharing a sequencing lane. RNA tomography, TIVA, RNA-capture arrays, and other NGS-based methods typically sacrifice the sampling resolution for the sequencing depth. In general, biological patterns or structures that occur reproducibly, such as in embryos, stable cell types, or clinically defined histologic features, benefit most from the NGS-dependent methods, as sampling of many identical cell types or structures will enable the statistically significant detection of small variations with higher

spatial resolution, whereas smFISH-based methods allow for quantifying the spatial variation in any small specimens, but mostly for validating a limited number of preselected genes.

ISSUES IN QUANTIFYING INTRINSIC AND EXTRINSIC VARIABLES IN CELLULAR DECISION-MAKING

In development, cells use information regarding lineage, intrinsic, and extrinsic factors to make decisions, and the genetic code specifies how such information is interpreted in different regions. The lineage information includes a historical record of master transcription factors and epigenetic modifications, whereas intrinsic factors include the cell state and the cell morphology. Extrinsic factors are composed of cell–cell or cell–extracellular matrix (ECM) interactions, such as the effect of stromal, endothelial, and inflammatory cells. The scale of cell–cell interactions can span few microns to many millimeters, implying that the quantification of extrinsic factors needs to occur at multiple scales.

The lineage information can be obtained by observing individual cells under a microscope;⁵⁴ however, this approach is limited to small and relatively transparent organisms, and they are low-throughput. Alternative approaches include combinatorial cell labeling using fluorescent proteins or somatic mutations for retrospective phylogenetic reconstruction.^{55,56} The former (i.e., Brainbow) lacks the combinatoric capacity to trace complete cell lineages, but it shows the cellular morphology, the cell–cell interaction, and the local tissue context.⁵⁵ In contrast, the latter (i.e., GESTALT) can be scaled up for complete phylogeny reconstruction of whole organisms;⁵⁶ however, it requires dissecting the whole organism into bulk tissues or single cells for NGS, destroying the spatial information. *In situ* DNA or RNA sequencing may provide a solution;^{18,22,29} however, sufficiently sensitive methods for sequencing multiple mutated loci in single cells *in situ* do not yet exist.

Intrinsic cellular variations can be measured using some of the existing single-cell analysis tools. For example, scRNA-seq can address the transcriptome heterogeneity, and the increasing throughput in combination with smFISH could make high-resolution cell-type mapping practical for well-defined developmental structures.⁵⁷ To put this into a perspective, however, a single 1-cm tissue mass contains up to a billion cells; therefore, scaling single-cell analyses to large tissues or organs may be impractical. A focused approach that combines labeled cells,

fluorescence-activated cell sorting (FACS), and scRNA-seq may better quantify cellular variations in defined cell populations.^{10,11} Of note, a variety of single-cell technologies, such as protein mass spectrometry, electrophysiology, and cellular perturbations, are all useful; however, many are relatively low-throughput compared with NGS-based single-cell methods, and they are more suited for addressing specific questions rather than whole-tissue mapping.

Unlike single-cell variations, the environmental context does not have a well-defined perimeter, as it can span multiple spatial scales. This is a well-known problem in functional genomics, where genetic features are visualized across a continuum of chromosomal window sizes and resolutions. It is also relevant in epidemiology, in which the emergence of nonrandom spatial patterns can alert one to a spreading epidemic, and the statistical significance is measured across a continuum of resolutions and scales.⁵⁸ Therefore, technological and statistical approaches to quantify nonrandom spatial patterns across various scales are necessary to quantify how genetic information is interpreted in a context-dependent manner *in vivo*.^{59–61}

Practically, *in vitro* experiments are useful for constraining extrinsic parameters, but they cannot simulate the native environment, especially across different spatial scales. On the other hand, direct sampling generates data that are dependent on the size and breadth of biopsies, and it introduces a significant amount of sampling noise at the lower end of measurement values. Pair-wise comparisons of intrinsic and extrinsic factors are incompatible with dissociated single-cell analysis, unless cells are labeled for reproducible sampling around individually marked cells. In contrast, *in situ* experiments can examine single cells and their native surroundings across a limited number of spatial scales; however, high-throughput molecular profiling *in situ* requires new types of technologies that require more development. One particular method that is potentially capable of co-localizing and quantifying lineage histories, intrinsic factors, and extrinsic forces across multiple tissue landscapes will now be the focus of the next discussion.

IN SITU SEQUENCING: KEY FEATURES, UNANSWERED QUESTIONS, AND CURRENT IMPLEMENTATION

High-throughput NGS has been used to read genetic barcode variants for lineage tracing, single-cell gene expression profiling, and microenvironmental gene expression profiling; however, it lacks the ability to co-localize such information onto the native tissue

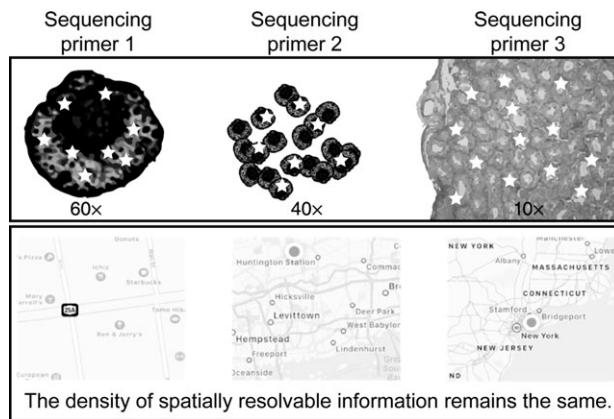


FIGURE 6 | Fluorescent *in situ* sequencing (FISSEQ) can utilize the programmable signal density to acquire a similar amount of information from multiplexed RNA detection regardless of the optical magnification or the transcript density.^{19,28} In this way, small and large biological patterns can be observed in the same specimen across a range of spatial scales, similar to how modern maps display the similar density of geographical information regardless of their resolution, making the map useful at all scales.

landscape. Because several popular NGS platforms sequence polymerase chain reaction amplicons on glass *in situ*,⁶² it was inevitable that some would attempt NGS inside single cells.¹⁹ Sequencing flow cells, however, are very different than the cellular environment, in terms of the subcellular transcript density, the tissue thickness, and the spatial scale at which imaging is typically done. The key aspect of FISSEQ is that it can change the random transcript sampling rate during 3D imaging to quantify the

gene expression heterogeneity across multiple scales and resolutions.^{19,28} To do so, FISSEQ amplicons are orders of magnitude brighter than smFISH signals, as low-magnification 3D microscopy makes single-molecule imaging demanding. Also, FISSEQ enables programmable subsampling using *in situ* sequencing primers so that individual transcript sequence can be resolved at any magnification (Figure 6).

To briefly summarize the FISSEQ workflow, which is detailed elsewhere,²⁸ fixed cells are immersed in an enzyme cocktail with random hexamers for reverse transcription *in situ*. The cDNA fragments are circularized and amplified using RCA, and sequencing-by-ligation is performed on a four-color confocal microscope (Figure 7). Base calling and sequence mapping against RefSeq are done for every pixel. The unmapped pixels are filtered out, and the mapped pixels are clustered using the sequence identity, allowing one to compare gene expression in multiple cellular regions, compartments, and morphologies. In the original publication,¹⁹ FISSEQ identified coding and non-coding RNAs associated with quiescent and proliferating fibroblasts, including those that regulated cellular migration and metabolism. In addition, it identified differential splicing of fibronectin *in situ* and detected a large shift in ribosomal RNA (rRNA) transcription in proliferating cells treated with EGF.

Interestingly, data from random hexamer-initiated *in situ* sequencing libraries are relatively scarce in major housekeeping genes despite their high abundance.^{19,28} Because FISSEQ amplicons are 100–1000 times larger than RNA–DNA duplexes in

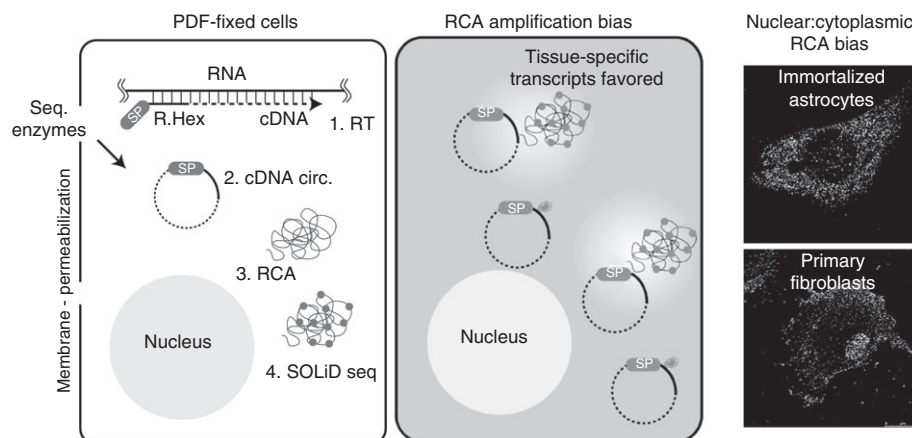


FIGURE 7 | Fluorescent *in situ* sequencing (FISSEQ) converts endogenous RNA molecules in fixed cells or tissues into short cDNA fragments *in situ* using random hexamer-primed reverse transcription (RT). (a) Each cDNA fragment contains a common sequencing adapter, which is then circularized prior to rolling circle amplification (RCA) *in situ*. RCA amplicons are then crosslinked to generate a stable 3D matrix of DNA molecules for *in situ* next-generation sequencing (NGS) reactions. FISSEQ then generates 3D images containing NGS reads at each pixel for data analysis.^{19,28} (b) Currently, the efficiency of RCA is not uniform across subcellular compartments, especially across different cell types.²⁸ We hypothesize that molecular crowding or liquid droplet phase transition may contribute to such observations; however, it is not clear whether such phenomenon is responsible for the relative paucity of housekeeping genes in FISSEQ.

smFISH, we suspect that molecular crowding or sequestration of particular transcripts impacts the RCA efficiency and imaging (Figure 7). Our ongoing experiments indicate that the efficiency of generating RCA amplicons from an smFISH probe is higher for the FISSEQ-compatible transcripts than for β -actin or tubulin mRNAs. Because of its potential implications for stress- or cell type-associated RNA compartmentalization, our group is investigating its biological basis in different cell and tissue types, including the differential subcellular cDNA amplification that are possibly indicative of liquid-phase transitions in the cytoplasm and the nucleus.⁶³

From a practical point-of-view, dedicated FISSEQ hardware is most cost-effective when using modular microscopy platforms capable of various 2D and 3D imaging modalities for examining subcellular regions, 3D organoids, and whole tissues at multiple resolutions and scales. While the existing protocols work well in most 2D or 3D cultured cells, RCA appears to be hindered in compact tissue sections due to the crowding effect, fixation, or embedding reagents. Limited tissue clearing may improve the RCA yield,^{44,45} but it is not obvious how each clearing method will affect the detection bias and sensitivity. For now, a dedicated high-speed spinning disk confocal microscope with four-color imaging appears to meet our needs (~\$350K, including customization). Our FISSEQ microscope from Nikon (Melville, NY) takes 1 min per 25 optical slices per field-of-view (FOV), and approximately 1000 FOVs per day of tiled imaging using PerfectFocus (~20 million mapped FISSEQ reads). With the installation of a perfusion set-up on the microscope stage for wash cycles, the hands-on time is less than 5 min per sequencing cycle.

HIGH-SENSITIVITY TARGETED *IN SITU* SEQUENCING: FUTURE PROSPECT

While FISSEQ enables unbiased transcriptome sampling in morphologically distinct subcellular and microenvironment regions, its estimated detection sensitivity is less than 0.005% compared to smFISH.²⁸ Yet, nonnormalized reads contain few major housekeeping genes from cultured cells¹⁹ and tissue sections (in preparation), even without depleting the rRNA.¹⁹ For example, common genes such as *ACTB*, *TUB1A1B*, *HSP90B1*, and ribosomal proteins (i.e., *RPS6*, *PRS23*, *RPS12*) are significantly underrepresented in FISSEQ versus RNA-seq compared with cell type-specific transcripts (odds ratio = 0.002–0.02; P -value < 10^{-12}).¹⁹ We hypothesize that this finding

could be useful for investigating RNA compartmentalization or localization, and we are currently extending our findings to multiple different cell and tissue types (Figure 7). Regardless, this causes targeted RNA-seq *in situ* to behave unpredictably depending on the cell or tissue type. Combined with the low detection sensitivity, the unpredictable or unknown nature of the FISSEQ detection bias makes selective sequencing of posttranscriptional modifications, mutations, indels, barcodes, and transcriptional reporters inefficient and prone to false negatives, especially in tissue sections. This means that investigating cell lineage, neural connections, and signaling reporters using *in situ* sequencing could reflect the technical bias rather than the biological variation. While tissue-clearing methods may reduce the detection heterogeneity,^{44,45} it could also worsen the detection sensitivity due to the prolonged nature of tissue clearing steps.

To overcome these limitations, smFISH-like approaches are needed, in which small oligonucleotides are bound directly to the RNA template for single-molecule imaging; however, it requires single-nucleotide resolution for sequential sequencing reactions *in situ*. Our laboratory is developing Heuristic *In Situ* Targeted Oligopaint sequencing (HISTO-seq), in which sequencing-by-ligation occurs directly on targeted RNA molecules *in situ* with high sensitivity and specificity using saturation sequencing enzyme kinetics (in progress). By incorporating the oligonucleotide cleavage and re-ligation chemistry from SOLiD (ThermoFisher), we are now determining the maximum HISTO-seq read length *in situ*, which is especially important for cell and lineage barcoding applications. We are also developing an approach to remove nonspecific probe binding, so that molecular quantification can occur in the absence of single-molecule imaging. By using programmable detection primers, targeted multiplex RNA-seq *in situ* with the smFISH-like sensitivity may now be possible.

To summarize, FISSEQ is useful for finding novel biomarkers and quantifying extrinsic factors in the environmental landscape in an unbiased manner; however, it lacks the sensitivity and the reproducibility for targeted RNA detection and genetic barcode sequencing in single cells. Moreover, much remains unknown about why FISSEQ detects predominantly non-housekeeping genes, emphasizing the need to see whether this extends to other cell and tissue types. Technologies like HISTO-seq require knowing the target genes-of-interest; however, if their high sensitivity and ease of use can parallel that of smFISH, they can become a powerful tool for investigating the single-cell intrinsic variation and cell lineage

information with subcellular, single-cell, and single-base resolution *in situ*.

BUILDING TOOLS TO INVESTIGATE HOW CELL KNOW WHERE THEY ARE

One of the foundational ideas in biology is that of positional information in normal development.⁶⁴ Turing first coined the term ‘morphogens’ and predicted that a simple regulatory network of diffusible molecules can generate robust patterns, suggesting how gene networks may encode patterns and forms in development. Wolpert then proposed the ‘French Flag’ model, in which the morphogen concentration gradient was interpreted differently through multiple signaling thresholds.⁶⁵ While they remain some of the most fundamental ideas in biology today, testing these hypotheses *in vivo* has proven difficult, except in a limited number of model organisms and isolated experimental systems. Regardless, the key assumption in these idealized models is that the diffusion rate is consistent across various tissue landscapes and that all responding cells interpret a particular morphogen concentration in the same way.

Despite many pioneers in the field, however, it is still not clear how cells interpret positional information to make decisions about its growth and differentiation.⁶⁶ If morphogen is one of the extrinsic factors, is the morphogen concentration the only variable in position-dependent cell fate specification? Do intrinsic cellular variations such as the cell cycle state,⁶⁷ biophysical forces,⁶⁸ morphology,⁶⁹ and cell–cell interactions play a role? If these decisions are also dependent on their cellular memory or lineage information that is not directly observable, how does one quantify the effect of intrinsic, extrinsic, and lineage information in cellular decision-making? (Figure 8).

The location awareness of cells is critical to understanding the origin of cancer as well.⁷⁰ The vast majority of cancers are derived from the epithelial cells involved in developmental patterning.⁷¹ Normally, such cells proliferate and differentiate progressively within a defined zone, and the cells that become lost commit apoptosis.⁷² In cancer-prone conditions, a disproportionate number of cells become lost and escape the geo-location fences set-up around the epithelial layer.⁷³ Therefore, investigating how cancer cells lose the ability to choose their fate based on their location may require examining intrinsic, extrinsic, and lineage information with single-cell resolution across various tissue regions.

Our collaborators and we are developing tools to sequence complete cell lineage information, similar to GESTALT⁵⁶ but *in situ*, using HISTO-seq, and we are also using it to classify single-cell transcriptional, morphological, and behavioral variations. Such information is then superimposed on the FISSEQ data to reveal how the landscape heterogeneity affects cellular decision-making. The most important benefit of *in situ* sequencing is that one can develop methods to capture other types of genetic data, including the somatic variation, the methylation signature, the DNA accessibility, and the chromatin topology. Rather than using bulk sampling-based or context-unaware technologies, adding the spatial dimension to these sequencing approaches could have a dramatic impact on our ability to select functionally relevant information.

CONCLUSION

The take-home message is that methods that depend on standard NGS (i.e., scRNA-seq, RNA tomography, LCM RNA-seq, spatial array capture RNA-seq) have limited spatial resolution because higher the sampling rate (i.e., single cells, ultra-thin tissue sections, microdissected regions) larger the random sampling noise for less abundant transcripts, requiring pooling of data from multiple individual samples. Also, the number of sampling points increases as the specimen size increases, making the economics of high-resolution sampling unfavorable. While shallow sequencing can provide some relief, it misses many low abundance and tissue-specific genes, capping its ability to spatially discriminate subtle cell type- or state-specific variations. These considerations make determining cell–cell and cell–microenvironment interactions challenging using NGS-based methods.

On the other hand, multiplexed smFISH depends on single-molecule imaging to discriminate signal from noise, hampering its application to whole-tissue imaging using low-magnification objectives. Because of its high sensitivity and nonspecific background probe binding, the lack of single-molecule imaging can lead to frequent false positives. Multiplexing approaches using single-molecule barcoding face a similar set of issues, as it is restricted by the optical diffraction limit, the practical voxel size, and the imaging time. Given these considerations, quantitatively surveying large tissues using multiplexed smFISH will likely be limited to a relatively small number of less abundant genes.

While FISSEQ is promising for investigating cell–cell and cell–microenvironmental interactions *in*

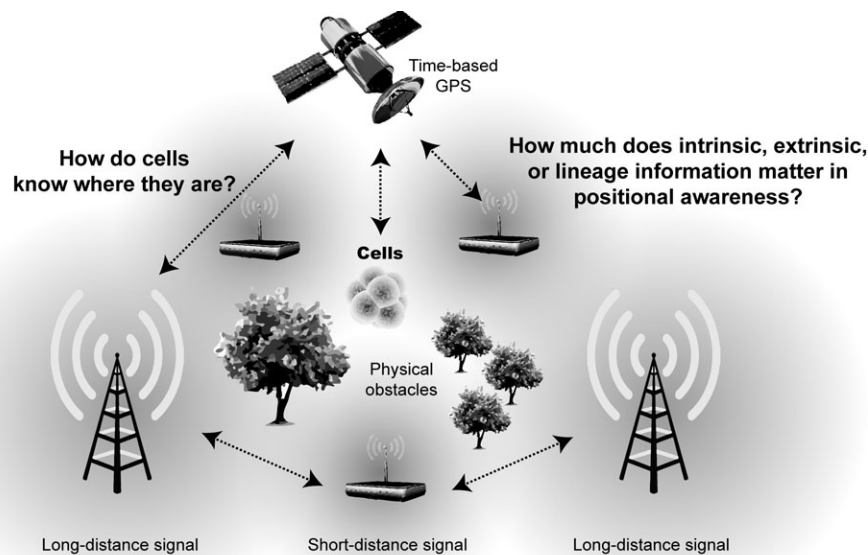


FIGURE 8 | Focusing technology development around a central question enables multiple creative approaches necessary to measure specific elements, rather than simply scaling up existing technologies. For example, positional information in developmental biology has been investigated under the assumption of idealized morphogen gradients without considering cellular and environmental variations. If true, any fluctuations in the signal strength due to environmental factors can make precise tissue patterning difficult over a global scale. To understand what makes interpreting positional information robust, our laboratory is developing three distinct *in situ* sequencing methods capable of measuring single-cell variations, microenvironmental heterogeneity, and cell lineages. In the future, it should be possible to perform selective knockdown of genetic pathways or optogenetic induction of morphogen signaling *in vivo* and quantify how intrinsic, extrinsic, and lineage-specific factors drive location-specific cell fate commitment and differentiation.

situ, it is too premature to say whether the method will be practical or informative in a wide range of cell and tissue types. FISSEQ on cultured cells, organoids, and whole-mount embryos are straightforward; however, a protocol for a wide range of fixed tissues is not yet available. The most interesting aspect of FISSEQ is its ability to selectively detect cell type- and state-specific transcripts without any human intervention, suggesting that the subcellular organization of RNA molecules could have functional implications for cell type- or tissue-specific transcripts.

Currently, the low sensitivity of FISSEQ limits its application for the targeted detection of specific RNA transcripts, including those bearing genetic barcode sequences. For example, it is typical for ~5000 specific transcripts to yield ~10 targeted FISSEQ amplicons per cultured cell, resulting in a high false negative rate due to stochastic biological or technical variations. This requires developing a targeted RNA sequencing method with the smFISH-like sensitivity with single-base resolution such as HISTO-seq. If successful, such approaches could allow for sequencing of various barcode associated with cell lineage tracing, signaling pathways, promoter activities, and Cas9-targeted gene perturbations *in situ*. In addition,

it could allow for tracking RNA processing, transport, and compartmentalization as well as discriminating multiple odorant receptors and their connectivity based on the transcript sequence.

Finally, finding a way out of the maze of new technologies can be challenging. The real-world information regarding the sensitivity, the specificity, the usability, the economics, and the availability require some digging, but they are generally available; however, a self-critical assessment about whether such technologies will actually address important biological questions are harder to find. For example, given that complex traits and cancer clonal evolution involve spatial mosaics of gene expression and somatic mutations, will cheaper sequencing of a large patient population be sufficient to map causative genetic elements? The current crop of sequencing technologies uses the abundance as a functional significance metric, which favors highly expressed transcripts over less abundant, less conserved, but often more tissue-specific noncoding RNAs. Will scaling up single-cell sequencing lead to the better understanding of the genome regulation, especially since ultra high-throughput single-cell approaches focus on the less noisy and more abundant transcripts?

The cutting-edge technology development can be as important as basic biological research because they can both lead to surprising and unexpected outcomes that can revolutionize science. Given the financial and opportunity cost of technology development and large-scale studies using such tools, however, the broad research community faces an important challenge of balancing the call to scale up the latest technologies versus the need to find the right technologies capable of addressing key biological questions. Because identifying fundamental biological questions

and variables that need to be measured are important elements in new technology paradigms, we advocate training young scientists familiar with technology development to ask fundamental biological questions, and those comfortable with rigorous experimental approaches to learn cutting-edge technology development. In the long run, we believe that such scientists may be best prepared to utilize emerging technologies, recognize their deficiencies, and build tools necessary to tackle many of the unaddressed questions in biology.

REFERENCES

1. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* 2013, 14:618–630.
2. Crosetto N, Bienko M, van Oudenaarden A. Spatially resolved transcriptomics and beyond. *Nat Rev Genet* 2015, 16:57–66.
3. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 2014, 509:371–375.
4. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015, 33:495–502.
5. Achim K, Pettit JB, Saraiva LR, Gavriouchkina D, Larsson T, Arendt D, Marioni JC. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* 2015, 33:503–509.
6. Levisky JM, Shenoy SM, Pezo RC, Singer RH. Single-cell gene expression profiling. *Science* 2002, 297:836–840.
7. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 2008, 5:877–879.
8. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* 2014, 11:360–361.
9. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015, 348:aaa6090.
10. Molyneaux BJ, Goff LA, Brettler AC, Chen HH, Brown JR, Hrvatin S, Rinn JL, Arlotta P. DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* 2015, 85:275–288.
11. Thomsen ER, Mich JK, Yao Z, Hodge RD, Doyle AM, Jang S, Shehata SI, Nelson AM, Shapovalova NV, Levi BP, et al. Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat Methods* 2016, 13:87–93.
12. Lovatt D, Ruble BK, Lee J, Dueck H, Kim TK, Fisher S, Francis C, Spaethling JM, Wolf JA, Grady MS, et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nat Methods* 2014, 11:190–196.
13. Junker JP, Noel ES, Guryev V, Peterson KA, Shah G, Huiskens J, McMahon AP, Berezikov E, Bakkers J, van Oudenaarden A. Genome-wide RNA tomography in the zebrafish embryo. *Cell* 2014, 159:662–675.
14. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, et al. Tumour evolution inferred by single-cell sequencing. *Nature* 2011, 472:90–94.
15. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebukova I, Loring JF, Laurent LC, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* 2012, 30:777–782.
16. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* 2012, 2:666–673.
17. Larsson C, Grundberg I, Söderberg O, Nilsson M. In situ detection and genotyping of individual mRNA molecules. *Nat Methods* 2010, 7:395–397.
18. Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wahlby C, Nilsson M. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* 2013, 10:857–860.
19. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SSF, Li C, Amamoto R, et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* 2014, 343:1360–1363.

20. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005, 309:1728–1732.
21. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010, 327:78–81.
22. Larsson C, Koch J, Nygren A, Janssen G, Raap AK, Landegren U, Nilsson M. In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nat Methods* 2004, 1:227–232.
23. Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, LeProust EM, Peck BJ, Emig CJ, Dahl F, et al. Multiplex amplification of large sets of human exons. *Nat Methods* 2007, 4:931–936.
24. Li JB, Levanon EY, Yoon J-K, Aach J, Xie B, LeProust E, Zhang K, Gao Y, Church GM. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* 2009, 324:1210–1213.
25. Ball MP, Li JB, Gao Y, Lee J-H, Leproust EM, Park I-H, Xie B, Daley GQ, Church GM. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 2009, 27:361–368.
26. Lee JH, Park IH, Gao Y, Li JB, Li Z, Daley GQ, Zhang K, Church GM. A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet* 2009, 5:e1000718.
27. Zhang K, Li JB, Gao Y, Egli D, Xie B, Deng J, Li Z, Lee J-H, Aach J, Leproust EM, et al. Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat Methods* 2009, 6:613–618.
28. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS, Aach J, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* 2015, 10:442–458.
29. Kalhor R, Mali P, Church GM. Rapidly evolving homing CRISPR barcodes. *bioRxiv* 2016:055863. doi: <http://dx.doi.org/10.1101/055863>.
30. Itzkovitz S, van Oudenaarden A. Validating transcripts with probes and imaging technology. *Nat Methods* 2011, 8:S12–S19.
31. Coskun AF, Cai L. Dense transcript profiling in single cells by image correlation decoding. *Nat Methods* 2016, 13:657–660.
32. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. *Science* 1998, 280:585–590.
33. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. *Nature* 2010, 463:913–918.
34. Raj A, van Oudenaarden A. Nature, nurture or chance: stochastic gene expression and its consequences. *Cell* 2008, 135:216–226.
35. Huang B, Babcock H, Zhuang X. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* 2010, 143:1047–1058.
36. Bates M, Huang B, Dempsey GT, Zhuang X. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science* 2007, 317:1749–1753.
37. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 2006, 3:793–795.
38. Lubeck E, Cai L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods* 2012, 9:743–748.
39. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 2008, 26:317–325.
40. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, Linnarsson S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* 2011, 21:1160–1167.
41. Chung K, Wallace J, Kim SY, Kalyanasundaram S, Andalman AS, Davidson TJ, Mirzabekov JJ, Zalocusky KA, Mattis J, Denisin AK, et al. Structural and molecular interrogation of intact biological systems. *Nature* 2013, 497:332–337.
42. Seo J, Choe M, Kim SY. Clearing and labeling techniques for large-scale biological tissues. *Mol Cells* 2016, 39:439–446.
43. Sylwestrak EL, Rajasethupathy P, Wright MA, Jaffe A, Deisseroth K. Multiplexed intact-tissue transcriptional analysis at cellular resolution. *Cell* 2016, 164:792–804.
44. Shah S, Lubeck E, Schwarzkopf M, He TF, Greenbaum A, Sohn CH, Lignell A, Choi HM, Gradinaru V, Pierce NA, et al. Single-molecule RNA detection at depth via hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* 2016, 143:2862–2867.
45. Chen F, Wassie AT, Cote AJ, Sinha A, Alon S, Asano S, Daugharthy ER, Chang JB, Marblestone A, Church GM, et al. Nanoscale imaging of RNA with expansion microscopy. *Nat Methods* 2016, 13:679–684.
46. Ku T, Swaney J, Park JY, Albanese A, Murray E, Cho JH, Park YG, Mangena V, Chen J, Chung K. Multiplexed and scalable super-resolution imaging of three-dimensional protein localization in size-adjustable tissues. *Nat Biotechnol* 2016, 34:973–981.

47. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012, 22:1775–1789.
48. Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, Lennartsson A, Ronnerblad M, Hrydziuszko O, Vitezic M, et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 2015, 347:1010–1014.
49. Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, Endy D, Church GM. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc Natl Acad Sci USA* 2013, 110:14024–14029.
50. Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 2015, 160:1246–1260.
51. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014, 343:84–87.
52. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016, 353:78–82.
53. Grun D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nat Methods* 2014, 11:637–640.
54. Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol* 1977, 56:110–156.
55. Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 2007, 450:56–62.
56. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 2016, 353:aaf7907.
57. Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015, 161:1202–1214.
58. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC; 2004.
59. Wahlby C. The quest for multiplexed spatially resolved transcriptional profiling. *Nat Methods* 2016, 13:623–624.
60. Mignardi M, Ishaq O, Qian X, Wahlby C. Bridging histology and bioinformatics—computational analysis of spatially resolved transcriptomics. *Proceedings of the IEEE* 99; 2016, 1–12.
61. Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 1989, 11:674–693.
62. Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet* 2010, 11:31–46.
63. Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* 2012, 149:768–779.
64. Green JB, Sharpe J. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development* 2015, 142:1203–1211.
65. Wolpert L. Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* 1969, 25:1–47.
66. Ashe HL, Briscoe J. The interpretation of morphogen gradients. *Development* 2006, 133:385–394.
67. Pauklin S, Vallier L. The cell-cycle state of stem cells determines cell fate propensity. *Cell* 2013, 155:135–147.
68. Bosveld F, Bonnet I, Guirao B, Tlili S, Wang Z, Petitalot A, Marchand R, Bardet PL, Marcq P, Graner F, et al. Mechanical control of morphogenesis by Fat/Dachsous/Four-jointed planar cell polarity pathway. *Science* 2012, 336:724–727.
69. Bosveld F, Markova O, Guirao B, Martin C, Wang Z, Pierre A, Balakireva M, Gague I, Ainslie A, Christophorou N, et al. Epithelial tricellular junctions act as interphase cell shape sensors to orient mitosis. *Nature* 2016, 530:495–498.
70. Bissell MJ, Labarge MA. Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment? *Cancer Cell* 2005, 7:17–23.
71. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science* 2013, 339:1546–1558.
72. Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. *Cell* 2011, 147:275–292.
73. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. *Nat Rev Cancer* 2009, 9:239–252.