Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues

Je Hyuk Lee[1,6], Evan R Daugharthy[1–3,6], Jonathan Scheiman[1,2], Reza Kalhor[2], Thomas C Ferrante[1], Richard Terry[1], Brian M Turczyk[1], Joyce L Yang[2], Ho Suk Lee[4], John Aach[2], Kun Zhang[5] & George M Church[1,2]

[1]Wyss Institute, Harvard Medical School, Boston, Massachusetts, USA. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [3]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. [4]Department of Electrical and Computer Engineering, University of California San Diego, California, USA. [5]Department of Bioengineering, University of California San Diego, La Jolla, California, USA. [6]These authors contributed equally to this work. Correspondence should be addressed to J.H.L. (jehyuklee@mac.com) or G.M.C. (gchurch@genetics.med.harvard.edu).

Published online 12 February 2015; doi:10.1038/nprot.2014.191

RNA-sequencing (RNA-seq) measures the quantitative change in gene expression over the whole transcriptome, but it lacks spatial context. In contrast, *in situ* hybridization provides the location of gene expression, but only for a small number of genes. Here we detail a protocol for genome-wide profiling of gene expression *in situ* in fixed cells and tissues, in which RNA is converted into cross-linked cDNA amplicons and sequenced manually on a confocal microscope. Unlike traditional RNA-seq, our method enriches for context-specific transcripts over housekeeping and/or structural RNA, and it preserves the tissue architecture for RNA localization studies. Our protocol is written for researchers experienced in cell microscopy with minimal computing skills. Library construction and sequencing can be completed within 14 d, with image analysis requiring an additional 2 d.

© 2015 Nature America, Inc. All rights reserved.

INTRODUCTION

Background

Cell type and function in tissues can be inferred from RNA or protein markers[1,2], but this approach to functional classification requires well-characterized biomarkers. Ideally, it would be preferable to define cell or tissue types using high-throughput molecular profiling *in situ* with high-resolution imaging. Indeed, several studies have surveyed global gene expression *in situ*, in which hundreds of organ tissue slices from multiple animals were individually interrogated using gene-specific probes[3–6]; however, such approaches represent a massive experimental undertaking, and they produce only an average view of tissue-specific gene expression.

In theory, multiplexed *in situ* RNA detection demands fewer samples, but so far this approach is limited by the number of spectrally distinct fluorophores and the optical diffraction limit of microscopy[7–11]. Alternatively, padlock probes[12–16] can capture specific RNA sequences from dozens of genes in parallel for targeted sequencing *in situ*[12]; however, padlock probes can have a substantial amount of probe-specific bias[17], and the approach cannot easily be scaled to the transcriptome. Given these challenges, *in situ* RNA profiling is typically restricted to a small number of well-annotated genes, and they can miss differences arising from unexpected signaling pathways or noncoding RNAs. In contrast, we wanted to develop an unbiased and transcriptome-wide sampling method for quantitative visualization of RNA *in situ*, preferably using direct molecular sequencing[18,19] for the detection of tissue-specific gene expression, RNA splicing and post-transcriptional modifications while preserving their spatial context; we call our method fluorescence *in situ* sequencing of RNA (FISSEQ).

Overview of the FISSEQ procedure

FISSEQ begins with fixing cells on a glass slide and performing reverse transcription (RT) *in situ*. After RT, the residual RNA is degraded to prevent it from competitively inhibiting CircLigase, and cDNA fragments are circularized at 60 °C. To prevent cDNA fragments from diffusing away, primary amines are incorporated into cDNA fragments during RT via aminoallyl-dUTP, and the primary amines are then cross-linked using BS(PEG)9. Each cDNA circle is linearly amplified using rolling-circle amplification (RCA) into a single molecule containing multiple copies of the original cDNA sequence, and the amine-modified RCA amplicons are cross-linked to create a highly porous and 3D nucleic acid matrix inside the cell (**Fig. 1a**).

In SOLiD sequencing by ligation, crucial enzymatic steps can be performed directly on a standard microscope at room temperature (25 °C). First, a sequencing primer is hybridized to multiple copies of the adapter sequence in RCA amplicons, followed by ligation of dinucleotide-specific fluorescent oligonucleotides. After imaging, the fluorophores are cleaved from the ligation complex, and ligation of fluorescent oligonucleotides is repeated six more times to interrogate dinucleotide pairs at every fifth position (**Fig. 1b**). To fill in the gaps between dinucleotide pairs, the whole ligation complex is stripped off, and four additional sequencing primers with a single base offset are used to repeat dinucleotide interrogation starting from positions *N*–1, *N*–2, *N*–3 and *N*–4, generating up to 35 raw 3D image stacks representing dinucleotide compositions at all base positions over time.

The raw images are enhanced using standard 3D deconvolution techniques to reduce the background noise, and our freely available MATLAB script performs image alignment to produce TIFF images that are then used for base calling using a separate python script (http://arep.med.harvard.edu/FISSEQ_Nature_Protocols_2014/). The base calls from individual pixels are then aligned to the reference transcriptome using Bowtie, and neighboring pixels with highly similar sequences are grouped into a single object generating a consensus sequence. The final data set includes the number of individual pixels per object, gene ID, consensus sequence, *x* and *y* centroid positions, number of mismatches, base call quality and alignment quality.

One of the key considerations early in the development of FISSEQ was imaging. Biological patterns, including RNA localization, occur

10x Genomics EX2079
Bruker v. Harvard
IPR2024-00526

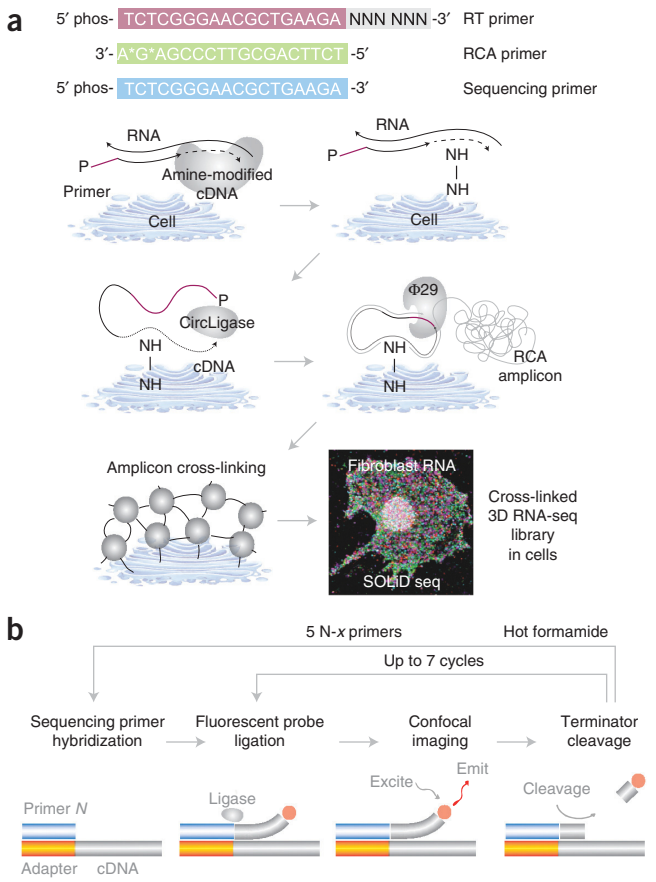

Figure 1 | Schematic overview of FISSEQ library construction and sequencing. (**a**) Fixed cells or tissues are permeabilized and reverse-transcribed *in situ* in the presence of aminoallyl-dUTP and adapter sequence-tagged random hexamers. The cDNA fragments are fixed to the cellular protein matrix using a nonreversible amine cross-linker and circularized after degrading the RNA. The circular templates are amplified using RCA primers complementary to the adapter sequence in the presence of aminoallyl-dUTP and stably cross-linked. The nucleic acid amplicons in cells are then ready for sequencing and imaging (fibroblast shown). (**b**) Each amplicon contains numerous tandem copies of the cDNA template and adapter sequence. A sequencing primer hybridizes to the adapter sequences in individual amplicons, and fluorescent eight-base probes interrogate the adjacent dinucleotide pair. After imaging, the three bases attached to a fluorophore are cleaved, generating a phosphorylated 5′ end at the ligation complex suitable for additional ligation cycles interrogating every fifth dinucleotide pairs. The whole process is repeated using four other sequencing primers with an offset to interrogate intervening base positions.

in a scale-dependent manner, in which some patterns are visible at one scale but disappear at another. Therefore, we developed our sequencing method specifically for confocal microscopy using a wide range of objectives, magnification, numerical apertures (NAs), scanning speed and depth. In addition, autofluorescence, cell debris and background noise are common in cell imaging, unlike in standard next-generation sequencing. Therefore, we developed an approach to classify individual pixels on the basis of their specific color transitions to detect true signals even in the noisy and/or low-intensity environment. Finally, we also developed a way to control the imaging density of single molecules, which enables the sequencing of a large number of molecules in single cells regardless of the microscopy resolution.

Comparisons with single-cell RNA-seq

More than one million mRNA reads per cell can be obtained from a single-cell RNA-seq experiment[20], but typically <100,000 reads per cell are from unique cDNA fragments, and PCR amplification accounts for the remainder[20–22]. In one study, the detection sensitivity of single-cell RNA-seq was estimated as ~10% or ~3% compared with single-molecule fluorescence *in situ* hybridization (FISH) or spiked-in controls, respectively[20]. This means that only ~300 genes are expected to have a coefficient of variation of <23% based on Poisson distribution; however, such genes are generally uninformative, and they include many housekeeping genes such as ribosomal subunit proteins (**Fig. 2a**), requiring that reads from multiple cells are combined to detect biologically meaningful gene expression differences between groups of single cells.

In FISSEQ, only ~200 mRNA reads per cell are obtained without rRNA depletion[23] (versus ~40,000 in single-cell RNA-seq); however, functionally important transcripts are enriched in FISSEQ by more than tenfold compared with single-cell RNA-seq (**Fig. 2b**). When examining a single spatial region of ~40 cells (~8,000 mRNA reads), the top-ranked genes lie substantially above the detection threshold, and they form highly reproducible cell type–specific annotation clusters[23]. Because of the relative absence of housekeeping genes, the high correlation (Pearson's r >0.9) between biological replicates in FISSEQ is driven by cell type– and/or function-specific genes rather than housekeeping genes.

To attain truly single-cell gene expression profiling that is biologically meaningful, FISSEQ may require a read depth per cell that is ~40 times deeper (~8,000 amplicons per cell).

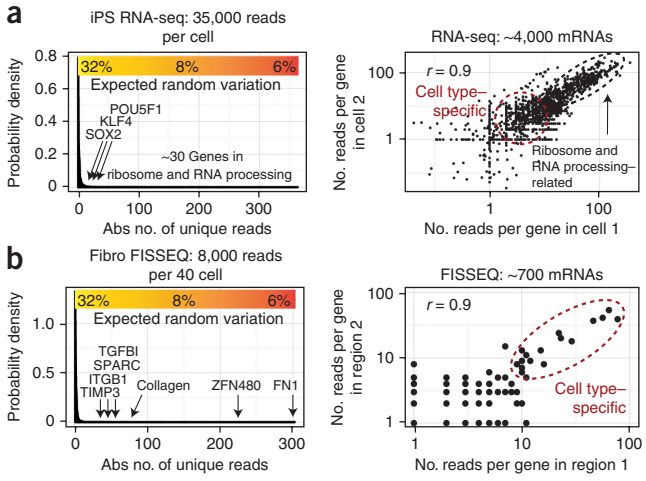

Figure 2 | Comparing single-cell RNA-seq with FISSEQ. (**a**) A typical single-cell RNA-seq[20] can generate more than one million reads per cell, but <10% represent unique reads from cDNAs, and they are composed largely of structural and/or housekeeping genes (i.e., ribosome-related). Many genes of interest are found near the detection limit with a large coefficient of variation, and the high correlation reported for single-cell RNA-seq is typically due to housekeeping genes. (**b**) The current version of FISSEQ combines mRNA reads from ~40 cells to obtain a comparable result, but the high correlation between biological replicates in FISSEQ results from mostly cell type–specific expression markers.

© 2015 Nature America, Inc. All rights reserved.

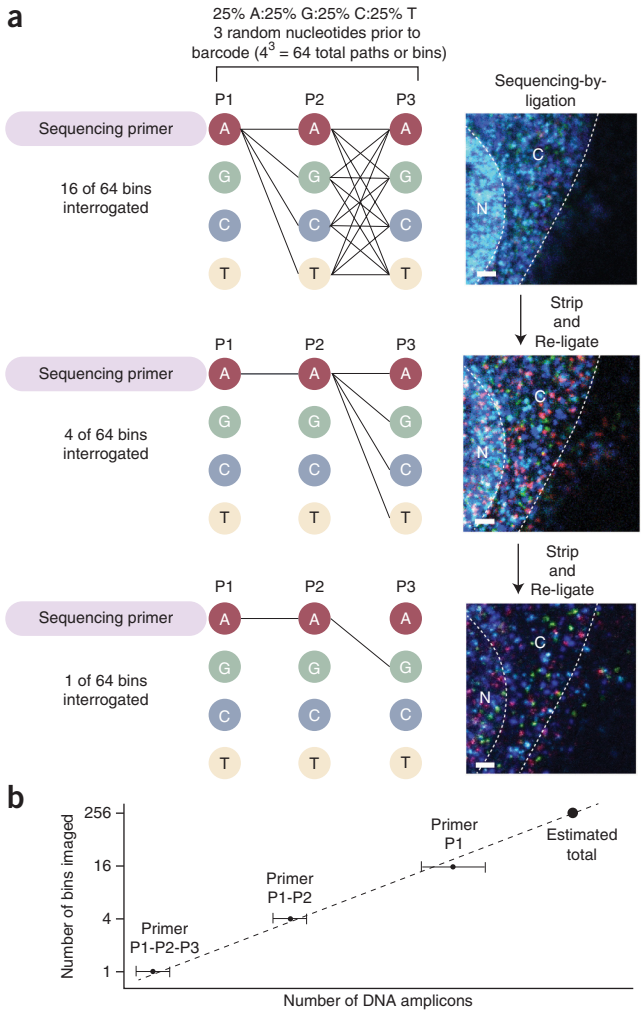

Figure 3 | Counting resolution-limited amplicons using partition sequencing. (**a**) The cDNA or padlock probe template can include three random nucleotides in equal proportions. By controlling the length of the complementary portion of the sequencing primer to the random bases, one can ligate fluorescent probes to different amplicon pools of varying sizes (fibroblasts; scale bars, 1 µm). This scheme works for single-base sequencing-by-ligation, and the SOLiD sequencing chemistry requires additional modifications to the bridge oligonucleotide. C, cytoplasm; N, nucleus. (**b**) Serial ligation reactions using the sequencing primers with 0–3 complementary bases to the random partitioning bases are analogous to doing a serial dilution experiment. The average count from each primer category can be used to extrapolate and estimate the actual amplicon count, regardless of the limitations in optical microscopy.

As the rRNAs comprise >80% of the reads in FISSEQ[23], it may be possible to increase the read depth by about fivefold by simply depleting rRNA *in situ*[24]. We expect another fivefold increase in the amplicon density by optimizing our reaction conditions, and a read depth of ~5,000 non-rRNA reads per cell may soon be possible. As individual amplicons of any density can be discriminated using partition sequencing[23] (**Fig. 3**), the actual size of each amplicon now becomes a limiting factor in the number of reads generated per cell.

Single-cell RNA-seq and FISSEQ are fundamentally limited by the efficiency of mRNA to cDNA conversion. In single-cell RNA-seq, this is estimated to be ~10% compared with single-molecule FISH[20], with a detection threshold of ~5–10 mRNA molecules per cell[21]. This means that most low-abundance genes are not detected in single-cell RNA-seq for a given cell. For FISSEQ, this value is harder to determine because not all genes are enriched in the same manner, but we estimate the current detection threshold at ~200–400 mRNA molecules per cell. After rRNA depletion and other improvements, the detection threshold may improve to ~10–20 mRNA molecules per cell; however, a large fraction of low-abundance genes will still remain undetected.

Comparisons with other approaches

Compared with microdissection[25,26] or photo-activated mRNA capture–based[27] single-cell RNA-seq[21,28–31], FISSEQ scales to large tissues more efficiently[32], and it can compare multiple RNA localization patterns in a nondestructive manner[23]. In addition, other methods require RNA isolation and PCR that can introduce a substantial amount of technical variability[20–22], assuming a Poisson distribution model of transcript abundance. In contrast, all samples can be processed together in a single well from cell culture to sequencing in FISSEQ.

Single-molecule FISH remains the gold standard for high-sensitivity detection of RNA in single cells[7–9,33–37]; however, spectral discrimination of hybridized probes can be difficult to multiplex, and it requires high-resolution microscopy. Recently, highly scalable FISH was demonstrated in single cells, in which sequential hybridization is used to barcode a color sequence for each transcript[10]. In theory, only seven hybridization cycles are required to interrogate 4^7 or >16,000 genes using four colors; however, this approach is limited by the sheer number of probes needed, and the optical diffraction limit prevents accurate quantification of highly abundant or aggregated transcripts.

The sensitivity of padlock probes is two orders of magnitude higher than FISSEQ for a given gene[12,13], but the use of locked nucleic acid makes this approach prohibitively expensive for multiplexing, and individual probes must be calibrated for measuring the relative RNA abundance. For certain applications, it may be possible to combine FISSEQ and padlock probes to interrogate a large number of loci *in situ*. In a recent study, sequencing was limited to short barcodes from dozens of gene-specific padlock probes[12], but now hundreds of thousands of padlock probes[17,38–41] can be discriminated using a 20-base barcode. In the same study, the microscopy resolution limited the number of targeted genes[12], but our partition sequencing[23] bypasses such limitations for highly multiplexed amplicon discrimination *in situ*.

Limitations

On a practical level, equipping a microscope for four-color imaging can currently cost up to $20,000 for a new filter set and a laser. Most users will need to reserve the microscope for 2–3 weeks so that sequencing can proceed uninterrupted. We have used laser-scanning confocal, wide-field epifluorescence and spinning-disk confocal microscopes and obtained comparable sequencing data that differ mainly in the read density. With the laser-scanning confocal microscope, imaging can take over 30 min per stack, but wide-field or spinning-disk confocal microscopes can image the same volume in 1–2 min. Reagent exchanges are done manually in the current protocol, but FISSEQ samples can remain on the microscope and be sequenced over 2–3 weeks.

© 2015 Nature America, Inc. All rights reserved.

On a technical level, a major limitation of our current protocol is the lack of rRNA depletion. Initially, we used rRNA as an internal control for library construction, sequencing and bioinformatics; however, this reduced the number of mRNA reads per cell. In primary fibroblasts, the rRNA reads comprised 40–80% of the total (ref. 23); therefore, if one were to deplete the rRNA[24], it might be possible to increase the number of mRNA reads per cell by about fivefold.

Another limitation is the lack of information on biases in our method. FISSEQ enriches for biologically active genes, enabling discrimination of cell type–specific processes with a small number of reads[23]; however, it is not clear how such enrichment occurs. We hypothesize that active RNA molecules are more accessible to FISSEQ, whereas RNA molecules involved in ribosome biogenesis, RNA splicing or heat-shock responses are trapped in ribonucleoproteins, spliceosomes or stress granules. It is now important to investigate and understand the molecular basis of such enrichment across multiple cell types and conditions and to correlate the result with the observed cellular phenotype.

Applications

The current FISSEQ protocol is suitable for most cultured cells and tissue sections, including formalin-fixed and paraffin-embedded (FFPE) tissue sections. Whole-mount *Drosophila* embryos, induced pluripotent stem cell (iPSC)-derived embryo bodies (EBs) and organoids are also compatible (**Table 1**). In FISSEQ, each sequencing read has a spatial coordinate, and the reads are binned according to the cellular morphology, subcellular location, protein localization or GFP fluorescence. A statistical test is then applied to identify enriched genes and pathways *de novo* and to discover possible biomarkers of the cellular phenotype[23]. This approach may be combined with padlock probes to detect evolving mutations and RNA biomarkers in cancers[12,13] or to compare gene expression in asymmetric cells or tissues.

FISSEQ may also sequence molecular barcodes in individual cells and transcripts, where expression or reporter (i.e., cDNA, promoter-GFP) libraries are examined in a pool of single cells for massively parallel functional assays and cell-lineage tracing. In essence, a practically unlimited number of DNA-associated cellular features may now be imaged, enumerated and analyzed across multiple spatial scales using the DNA sequence as a temporal barcode.

Experimental design

General considerations. This protocol details the method described in our original report[23], in which endogenous RNAs in cultured fibroblasts were sequenced on a confocal microscope. The availability of a microscope and computational resources will guide the general experimental approach (**Table 2**). We provide basic computational tools along with a sample data set, but a background in python, MATLAB, ImageJ and/or R is helpful for analyzing a large number of images. If such expertise is not available, we recommend focusing on a few regions of interest with well-demarcated features for comparing gene expression using our custom scripts[23]. After outlining the experiment, one should download our sample image, software and data set and become familiar with image and data analysis (http://arep.med.harvard.edu/FISSEQ_Nature_Protocols_2014/). One should then finalize the experimental design and define the imaging parameters (i.e., area, thickness, resolution and magnification).

Cell and tissue fixation. We have been able to fix and generate *in situ* sequencing libraries in a wide number of biological specimens (**Table 1**). The only case in which we failed was a hard piece of bone marrow embedded in Matrigel, which detached from the glass surface after several wash steps. Fixation artifacts can include changes in subcellular RNA localization, cell swelling, incomplete permeabilization and RNA leakage. Certain primary cell types are also sensitive to cold[42], whereas transformed cell lines or stem cells appear to be less sensitive (**Supplementary Fig. 1**). If you are using FISSEQ to study subcellular localization, we recommend fixing cells by adding warm formalin directly into the growth medium to a final concentration of 10% (vol/vol).

Cell and tissue sample mounting. For high-resolution imaging, we recommend poly lysine– or Matrigel-coated glass-bottom

© 2015 Nature America, Inc. All rights reserved.

TABLE 1 | Specimens tested for FISSEQ library construction.

Types	Fixation	Mounting substrate	Permeabilization	Notes
HeLa, 293A, COS1, U2OS, iPSC, primary fibroblasts and bipolar neurons	10% (vol/vol) formalin or 4% (vol/vol) PFA	Poly lysine–coated coverslip (Matrigel for iPSCs)	70% (vol/vol) ethanol or 0.25% (vol/vol) Triton X-100 (0.1 N HCl optional)	Changes in temperature can cause altered mRNA localization
Mouse embryo FFPE section (20 µm)	Already fixed	Superfrost Plus glass slide	0.1% (wt/vol) pepsin in 0.1 N HCl	Use silicone isolators (Grace Bio-Labs)
Mouse brain fresh-frozen section (20 µm)	10% (vol/vol) formalin	Poly lysine–coated coverslip	0.1% (wt/vol) pepsin in 0.1 N HCl	Use silicone isolators (Grace Bio-Labs)
iPS-derived 3D organoids	10% (vol/vol) formalin	Poly lysine–coated coverslip (embed in Matrigel and fix with 4% PFA)	0.25% (vol/vol) Triton X-100 and 0.1 N HCl	10% (vol/vol) formalin is less effective for fixing Matrigel
Dechorionated whole-mount *Drosophila* embryos	10% (vol/vol) formalin	Poly lysine–coated coverslip (embed in Matrigel and fix with 4% PFA)	100% (vol/vol) methanol then PBS with 0.2% (vol/vol) Triton X-100 and 0.2% (vol/vol) Tween-20	10% (vol/vol) formalin is less effective for fixing Matrigel

TABLE 2 | Comparison of the microscopy platforms tested for FISSEQ.

	Model	Pros	Cons	Uses
Wide-field epifluorescence	Nikon TE-2000	Fast imaging Simple setup	Poor axial resolution Low signal-to-noise ratio Lower read depth	Thin cells and tissue sections Whole-cell barcode labeling
Scanning confocal	Zeiss LSM 710 confocal Leica TCS SP5 Confocal	Good axial resolution Scanning zoom Flexible pixel density	Slow imaging	High-resolution FISSEQ of a single region
Spinning-disk confocal	Yokogawa CSU-W1	Fast imaging Good axial resolution	Fixed pixel density	All purpose

dishes, but 96-well plastic-bottom plates can be used for simple protocol optimization. Tissue sections can be mounted using a standard mounting procedure, and we advise inexperienced users to consult those who have experience in the art of tissue mounting. For nonadherent cell types and whole-mount specimens, we recommend fixing samples embedded in Matrigel using 4% (vol/vol) paraformaldehyde (PFA) on a glass-bottom dish.

RT *in situ*. The length of RT primers should be <25 bases to prevent self-circularization. We perform RT overnight for most samples, but 1 h is often sufficient for cell monolayers. A negative control without RT should be included to rule out self-circularization of the primer. A positive control primer with the adapter sequence plus a synthetic sequence (~30 additional bases) can be used to check RCA and imaging parameters. Other than the 5′ region of highly abundant mCherry transcripts[23], we have not had consistent results with targeted RT; we typically see very few amplicons regardless of the primer design. In contrast, random hexamers (24 bases) and poly-dT primers (33 bases) work well across all conditions. Some of the possible reasons for failure may include poor target accessibility and competitive inhibition of CircLigase by nonspecifically bound sequence-specific RT primers that are capable of self-circularization. Possible solutions include the use of locked nucleic acid (LNA)-based RT primers for high-temperature hybridization[13], ligation of the adapter sequence after RT and tiling multiple RT probes across a gene target. We have yet to try these alternatives.

Generation of amplicon matrix. Aminoallyl-dUTP is a dTTP analog commonly used in fluorescence labeling of cDNA[43], which we use for cross-linking nucleic acids; however, the efficiency of RT and RCA is inversely correlated with the concentration of aminoallyl-dUTP[23]. The cross-linker, bis(succinimidyl)-nona-(ethylene glycol) or BS(PEG)9, is functionalized with *N*-hydroxysuccinimide (NHS) ester groups at both ends[44], and it forms a stable covalent bond with primary amine groups provided by aminoallyl-dUTP at pH 7–9. The cross-linking density can be enhanced by increasing the concentration of aminoallyl-dUTP or BS(PEG)9, or by increasing the pH. Cross-linking after RT is optional, but cross-linking of RCA amplicons is essential for high-quality sequencing reads.

Sequencing. We use sequencing-by-ligation[18,19,45] (SOLiD[46,47]) because it works well at room temperature, and so a heated stage is not required. SOLiD uses a dinucleotide detection scheme in which a base position is interrogated twice per sequencing run[46,47], and this can reduce the base calling error rate; however, converting the color sequence to the base sequence is not straightforward because of its propensity to propagate errors, and sequence analysis must remain in the color space (**Box 1** and **Fig. 4**). In comparison, sequencing-by-synthesis (Illumina) works at 65 °C for primer extension and cleavage, and it uses proprietary fluorophores that require a heated flow-cell and a custom imaging setup. As sequencing-by-synthesis can generally yield a much longer read length, we are currently investigating its compatibility with FISSEQ.

Box 1 | SOLiD sequencing chemistry

The SOLiD sequencing chemistry consists of multiple reaction cycles in which a sequencing primer is extended using fluorescent eight-base probes via sequential DNA ligation. The fluorescent amplicons are then imaged, and the last three bases and the fluorophore are cleaved, followed by the ligation of another eight-base probe. These steps are repeated using four additional sliding primers to record the dinucleotide color values from starting positions 1-6-11-16-20-26-31 (primer *N*), 0-5-10-15-20-25-30 (primer *N*–1), 4-9-14-19-24-29-34 (primer *N*–2), 3-8-13-18-23-28-33 (primer *N*–3) and 2-7-12-16-22-27-32 (primer *N*–4; **Fig. 4a**). Most bases are represented by two sequential colors, and although each color represents up to four possible dinucleotide combinations the exact nucleotide sequence can be determined if the identity of any one base is known (i.e., the base identity in the sequencing primer). For example, AAGCAGTCA is equivalent to BORGOGOG (B: blue, O: orange, R: red and G: green; **Fig. 4b**); however, the conversion table alone cannot assign the base identity from color codes. However, if one base is known (i.e., first base is A in BORGOGOG), assigning the base identity is relatively straightforward (**Fig. 4c**). One disadvantage is that any missing or wrong base calls can affect the whole read, and it makes sequence-to-sequence comparisons impossible. Therefore, the SOLiD sequencing reads and the reference database must remain in the color space for sequence alignment, and the user should keep this in mind when designing a custom sequence analysis pipeline.

© 2015 Nature America, Inc. All rights reserved.

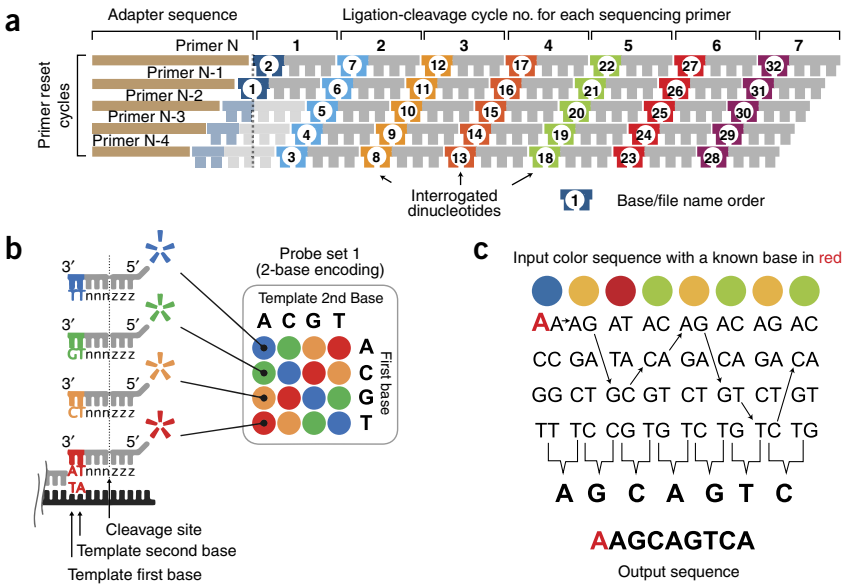

Figure 4 | Schematic overview of the SOLiD color-coding and decoding scheme. (**a**) The base position within the template sequence is enclosed by white circles and should be used for naming the image files, and the actual sequencing cycle numbers are noted on both sides. Each ligation extension is shown in different colors, and cycles 15, 22 and 29 are shown in gray, as no images are acquired for these cycles. The red box at cycle 8 denotes a known base identity. (**b**) SOLiD dinucleotide coding scheme. (**c**) SOLiD color space decoding scheme. As long as any one of the base identities are known (here in red), the color space sequence can be converted to the nucleotide sequence. Image reproduced from Life Technologies (ref. 47). © 2014 Thermo Fisher Scientific, Inc. Used under permission.

Partition sequencing. T4 DNA ligase has a single-base specificity at the ligation junction[18], and sequencing primers differing by one base can recognize different sets of amplicons[23]. By dividing imaging over multiple separate runs, spatially overlapping amplicons can be enumerated using multiple sequencing primers even on a low-resolution microscope; however, this requires full automation for the increased number of sequencing runs per sample. Without automation, partition sequencing is better suited for quantifying short barcode sequences rather than full RNA sequences *in situ*[12] (**Fig. 3**).

Imaging. Epifluorescence microscopy can generate a reasonable number of alignable reads from relatively thin specimens (<5 μm), such as HeLa cells[23], but thicker samples require confocal microscopy to obtain high-density reads. Spinning-disk confocal microscopy is markedly faster than laser-scanning confocal microscopy, and it has a good balance of imaging speed and axial resolution. An automated stage capable of finding a *z*-stack across multiple *x*-*y* tiles is highly desirable (**Table 2**).

In FISSEQ, individual amplicons can be detected using objectives with a NA of 0.4 or greater. The magnification required is determined by the biological question and the amplicon density[48]. Typically, we use a 20× NA 0.75 objective to examine tissue sections and cultured cell monolayers, whereas 40× NA 0.8 and 63× NA 1.2 water-immersion objectives are used for high-resolution imaging of single cells. We have observed noticeable chromatic aberration in our experiments, depending on the objectives used. The degree of chromatic aberration should be measured using image calibration beads (i.e., FocalCheck fluorescence microscope test slide) before sequencing, and they should be calibrated by the microscope vendor if necessary.

For each imaging setup, the user should determine the ideal Nyquist rate. This value can be calculated using http://www.svi.nl/NyquistCalculator. The *x*-*y* pixel and *z*-step sizes should not be >1.7 times the Nyquist value for image deconvolution. Four-color imaging should proceed from the longest to the shortest wavelength (i.e., Cy5, Texas Red, Cy3 and FAM), and an intensity histogram should be used to adjust the laser power to prevent saturated pixels. The intensity histogram should be consistent across fluorescence channels and sequencing cycles. To use our software, the image file name must be standardized: <Position>_<Primer #>_ <Ligation #>_<Date_Time>.extension (e.g., 06_N1_2_2013_10_25_11_57_18.czi).

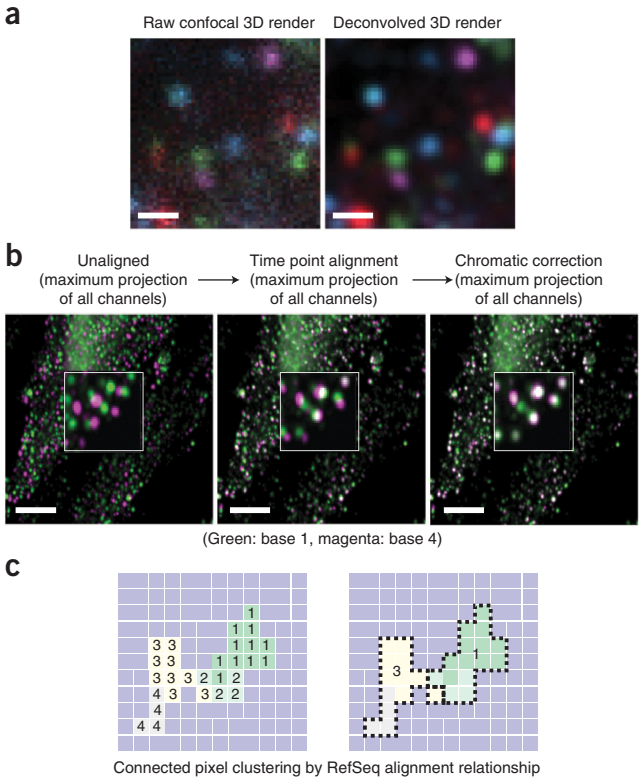

Figure 5 | Example of image analysis, registration and sequence clustering. (**a**) A four-color confocal image stack is deconvolved using a CLME algorithm with ten iterations and a signal-to-noise ratio of two (scale bars, 1 μm). (**b**) Sequencing images for base 1 and base 4 (left) are aligned using a composite channel across various time points (middle) and then using a composite time projection across various channels (right; scale bars, 5 μm). (**c**) Individual nonzero pixels are aligned to the reference sequence database (i.e., human RefSeq). Highly related sequences connected to the neighboring pixels are then grouped into a single cluster.

© 2015 Nature America, Inc. All rights reserved.

Image analysis tools. In practice, the extent of image processing and analysis is dictated by the available imaging tools and computing resources[49]. We use Bitplane Imaris for data visualization and movie creation and Scientific Volume Imaging (SVI) Huygens for 3D deconvolution. Although they are easy to use, scalable and relatively fast, their cost may be out of reach for small laboratories; however, free and/or open-source alternatives are also available[49–51].

Image deconvolution. We use 3D deconvolution[52] to reduce the out-of-focus background and to improve the quality of base calls (**Fig. 5a**). High-quality 3D deconvolution requires sampling near the Nyquist rate, but this increases the image acquisition and deconvolution time, as well as the file size. We generally recommend using high-quality confocal imaging and minimal 3D deconvolution for FISSEQ. The use of 3D deconvolution to compensate for low-quality imaging will not necessarily improve the quality or the number of sequencing reads. We provide a sample data set containing raw and deconvolved image stacks from a successful 30-base sequencing experiment for practice (http://arep.med.harvard.edu/FISSEQ_Nature_Protocols_2014/).

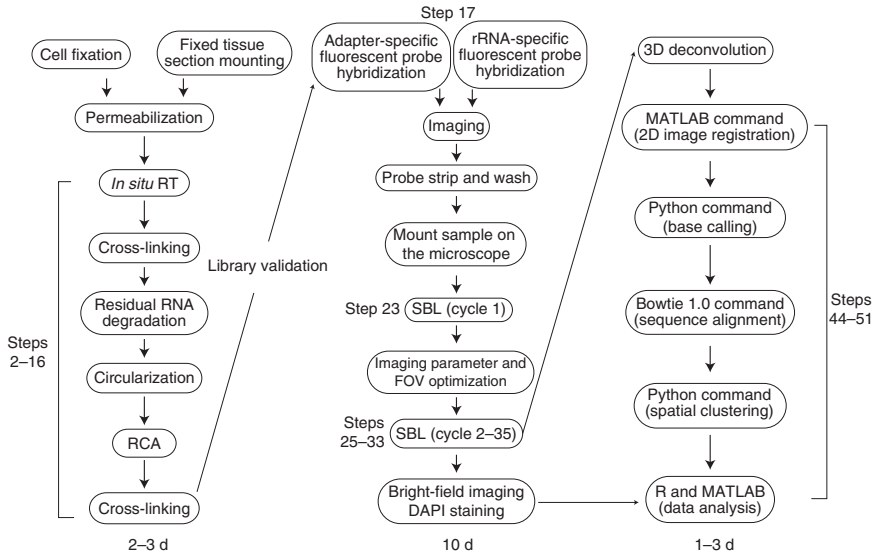

Figure 6 | Schematic overview of FISSEQ experimental and analysis steps. FOV, field of view; SBL, SOLiD sequencing-by-ligation.

Image registration, base calling and sequence alignment. As long as the input image files are correctly named, our software will generate the maximum intensity projection, register the images and correct for chromatic shifts[23] (**Fig. 5b**). The resulting images are used for base calling and sequence alignment to human RefSeq (**Fig. 5c**), but our software does not generate *z*-coordinates for sequencing reads, as it uses maximum intensity projection for base calling. We provide a sample data output and screen logs for troubleshooting our bioinformatics software (http://arep.med.harvard.edu/FISSEQ_Nature_Protocols_2014/).

Data analysis. Our software generates a tab delimited text file that contains 10,000–50,000 aligned reads per field of view. We recommend RStudio with the latest version of R installed for plotting reads by RNA classes, position, cluster size, quality, gene name, strand and so on. We provide a sample R session file that is used for FISSEQ data analysis as an introduction to statistical computing and for assessing the quality of FISSEQ data set (http://arep.med.harvard.edu/FISSEQ_Nature_Protocols_2014/).

Level of expertise required for the protocol. FISSEQ is at the intersection of cell imaging and functional genomics, and it has generated much interest from cell biologists who are not familiar with RNA-seq. Our protocol is aimed at such researchers, who are familiar with cell image analysis but have few computing skills (**Fig. 6**). FISSEQ library construction can be performed by anyone with basic molecular biology skills, but image acquisition is best done with help from an imaging core specialist for the initial setup. Once the equipment, software, imaging and deconvolution parameters are finalized, a capable technician, graduate student or post-doc can perform manual sequencing on a microscope with some training and practice. Image and sequence analysis using our software can be performed by anyone familiar with the Unix environment, but statistical data analysis requires either a graduate student or post-doc familiar with statistical tools and concepts.

Considerations about the laboratory facilities. All steps in FISSEQ library construction can be carried out in a standard laboratory setting. A vacuum line facilitates solution aspiration and reagent exchanges, and we do not find RNA degradation or PCR contamination to be a notable problem in our method. We advise having a dedicated microscope with proper excitation and emission filters on a vibration isolation table in a low-traffic area.

MATERIALS

REAGENTS

- Starting material of interest. The PROCEDURE is written for cultured cells in glass-bottom dishes or for tissue sections on glass-bottom dishes or coverslips. However, it can be adapted for use on a range of starting materials (**Table 1**)
- Acetone (for detaching a Petri dish glued to the microscope stage)
  ! CAUTION Acetone is highly flammable. Work in a well-ventilated area.
- Aminoallyl-dUTP, 4 mM (AnaSpec, cat. no. 83203)
- Betaine, 5 M (included in CircLigase II kit, Epicentre, cat. no. CL9025K)
- BS(PEG)9, 100 mg (Thermo Scientific, cat. no. 21582)
  ▲ CRITICAL BS(PEG)9 loses its effectiveness 1 month after reconstitution in DMSO. Prepare a fresh batch every month, especially if it has been frozen and thawed repeatedly.
- CircLigase II kit (Epicentre, cat. no. CL9025K)
- Cleave solution 1 (Applied Biosystems, cat. no. 4406489)
- Cleave solution 2.1 kit (Applied Biosystems, cat. no. 4445677)
  ! CAUTION Contains toxic organoamine. Wear gloves and work in a well-ventilated area.

© 2015 Nature America, Inc. All rights reserved.

- Cyanoacrylate adhesive, optical grade (VWR International, cat. no. 19806-00-1)
- Diethylpyrocarbonate (DEPC)-treated water (Santa Cruz Biotechnology, cat. no. sc-204391)
- DMSO (Sigma-Aldrich, cat. no. D8418)
- dNTP, 25 mM (Enzymatics, cat. no. N2050L)
- Ethanol, 70% (vol/vol) (in DEPC-treated water)
- Formalin, 10% (vol/vol) (Electron Microscopy Sciences, cat. no. 15740)
  ! CAUTION Wear gloves and work in a well-ventilated area. Dispose of waste per institutional guideline.
- Formamide (Sigma-Aldrich, cat. no. 221198)
- HCl, 0.1 N (in DEPC-treated water)
- Immersol W 2010 (ne = 1.33) for water-immersion lens (Carl Zeiss Microscopy, cat. no. 444969-0000-000)
- Instrument buffer, 10× (Applied Biosystems, cat. no. 4389784)
- Moloney-murine leukemia virus (M-MuLV) reverse transcriptase (Enzymatics, cat. no. P7040L)
- Mineral oil (Sigma-Aldrich, M5904)
- $MnCl_2$ (included in CircLigase II kit, Epicentre, cat. no. CL9025K)
- Nuclease-free water, not DEPC-treated (Life Technologies, cat. no. AM9932)
- Pepsin, 1 g (dissolve in 10 ml of H_2O and store it at −20 °C; Affymetrix, cat. no. 20010)
- ϕ29 DNA polymerase (Enzymatics, cat. no. P7020-HC-L)
- PBS (Life Technologies, cat. no. 10010023)
- RNase, DNase-free (Roche Applied Science, cat. no. 11579681001)
- RNase H (Enzymatics, cat. no. Y9220F)
- RNase inhibitor (Enzymatics, cat. no. P9240L)
- RNaseZap (Life Technologies, cat. no. P9780)
- Silicone isolator (Grace Bio-Labs, cat. no. 664304)
- Sodium acetate, 3 M (pH 7.5)
- SOLiD ToP sequencing kit fragment library F3 tag MM50 (Applied Biosystems, cat. no. 4449388)
- Saline sodium citrate (SSC), 20× (Roche Applied Science, cat. no. 11666681001)
- Streptavidin–Alexa Fluor 647 (Life Technologies, cat. no. S32357)
- T4 DNA ligase (Enzymatics, cat. no. L6030-LC-L)
- Tris solution, 1 M (G-Biosciences, cat. no. R002)
- Trisodium citrate dihydrate (Sigma-Aldrich, cat. no. C8532)
- Triton X-100, 10% (vol/vol) solution (Sigma-Aldrich, cat. no. 93443)

RT, RCA and sequencing primers (all are in 5′–3′ orientation)
- Random hexamer RT primer, 100 μM in nuclease-free H_2O (/5phos/TCTCGGGAACGCTGAAGANNNNNN; hand-mixed, Integrated Data Technologies (IDT))
- RCA primer, 100 μM (TCTTCAGCGTTCCCGA*G*A; * is phosphorothioate)
- Sequencing primer *N*: /5phos/TCTCGGGAACGCTGAAGA (HPLC purified)
- Sequencing primer *N*–1: /5phos/CTCGGGAACGCTGAAGA (HPLC purified)
- Sequencing primer *N*–2: /5phos/TCGGGAACGCTGAAGA (HPLC purified)
- Sequencing primer *N*–3: /5phos/CGGGAACGCTGAAGA (HPLC purified)
- Sequencing primer *N*–4: /5phos/GGGAACGCTGAAGA (HPLC purified)

Control primers (5′–3′)
- Adapter-specific probes, 100 μM (/56-FAM/TCTCGGGAACGCTGAAGA)
- Adapter-specific probes, 100 μM (/5TYE563/TCTCGGGAACGCTGAAGA)
- Adapter-specific probes, 100 μM (/5TEX615/TCTCGGGAACGCTGAAGA)
- Adapter-specific probes, 100 μM (/5TYE665TCTCGGGAACGCTGAAGA)
- 18S rRNA detection primer1: /5biotin/ GCTACTGGCAGGATCAACCAGGTA
- 18S rRNA detection primer2: /5biotin/ TACGCTATTGGAGCTGGAATTACC
- 18S rRNA detection primer3: /5biotin/ GTTGAGTCAAATTAAGCCGCAGGC
- 18S rRNA detection primer4: /5biotin/ TTGCAATCCCCGATCCCCATCACG
- 28S rRNA detection primer1: /5biotin/ CCACGTCTGATCTGAGGTCGCG
- 28S rRNA detection primer2: /5biotin/ CACGCCCTCTTGAACTCTCTCTTC
- 28S rRNA detection primer3: /5biotin/ CTCCACCAGAGTTTCCTCTGGCT
- 28S rRNA detection primer4: /5biotin/ TGAGTTGTTACACACTCCTTAGCG
- 28S rRNA detection primer5: /5biotin/ CGACCCAGCCCTTAGAGCCAATC
- 28S rRNA detection primer6: /5biotin/ GACAGTGGGAATCTCGTTCATCCA
- 28S rRNA detection primer7: /5biotin/ GCACATACACCAAATGTCTGAACC

EQUIPMENT
- 4 °C and −20 °C storage units
- Centrifuge for 1.5- and 2-ml tubes
- Dry-block heater for microtubes at 80 °C
- Falcon conical centrifuge tubes (15 and 50 ml; Fisher Scientific, cat. nos. 14-959-49B and 14-432-22)
- Flexible plastic i.v. catheter for reagent aspiration (Terumo, cat. no. SR*FF2419)
  ! CAUTION The catheter comes with a plastic outer sheath and a sharp needle in the middle. The needle must be carefully removed and discarded into a sharps container.
- FocalCheck fluorescence microscope test slide (Life Technologies, cat. no. F36909)
- Glass bottom MatTek dish (Poly lysine-treated: cat. no. P35GC-1.5-14-C, Poly lysine-treated 96-well plate: cat. no. P96GC-1.5-5-F)
- Glass Pasteur pipettes (autoclaved)
- Incubators at 30 °C, 37 °C (humidified) and 60 °C
- Inverted confocal microscope, PC and image acquisition software (see Equipment Setup)
- Microscope stage insert, metal (for securely gluing the specimen holder)
- Nonsterile syringes, 10 ml (BD Biosciences, cat. no. 301029)
- RNase-free microtubes (Eppendorf, cat. no. 0030 121.589)
- Sealable plastic container or Ziploc bags (for CircLigase reaction at 60 °C)
- Vacuum flask, trap and tubing

PC and software requirements
- Access to a high-performance computing cluster (remote host)
- Bowtie 1.0 or earlier (http://bowtie-bio.sourceforge.net) on the remote host
  ▲ CRITICAL Bowtie 2.0 or higher does not work with SOLiD sequencing.
- Fiji/ImageJ (http://fiji.sc/Fiji) on a PC
- MATLAB (http://www.mathworks.com) on the remote host
- Python 2.7 (https://www.enthought.com/products/canopy/) on the remote host **▲ CRITICAL** Other versions of python lack the required modules for running our script.
- R (http://www.r-project.org) and RStudio (http://www.rstudio.com) on a PC
- Windows PC or Mac with 16 GB RAM minimum
- Optional: SVI Huygens 3D deconvolution software (commercial), Bitplane Imaris 3D rendering software (commercial)

REAGENT SETUP

Triton X-100, 0.25% (vol/vol) Dilute 0.25 ml of 10% (vol/vol) Triton X-100 in DEPC-treated H_2O to a total volume of 10 ml. Store it at room temperature for up to 6 months.

SSC, 2× Dilute 20× SSC in H_2O and adjust the final volume to 50 ml. Store it at room temperature for up to 6 months.

SSC, 1× Dilute 20× SSC in H_2O and adjust the final volume to 50 ml. Store it at room temperature for up to 6 months.

SASC, 5× Make 0.75 M sodium acetate, 75 mM tri-sodium citrate, and then adjust the pH to 7.5 using acetic acid in H_2O to a final volume of 50 ml. Store it at room temperature for up to 6 months.

RCA primer hybridization buffer Dilute 20× SSC, 2× SASC and 30% (vol/vol) formamide in H_2O. Store the buffer at room temperature for up to 6 months.

Strip buffer Strip buffer is 80% (vol/vol) formamide in H_2O and 0.01% (vol/vol) Triton X-100 in a final volume of 50 ml. Store the buffer at room temperature for up to 6 months.

Cleave solution 2.1, reconstituted Mix 1 ml of cleave solution 2.1 Part 1 with 2.75 ml of cleave solution 2.1 Part 2. Store it at 4 °C in the dark for up to 24 h.

EQUIPMENT SETUP

Microscope setup Configure a four-channel microscope with appropriate excitation light sources and emission filters: FITC-488 excitation, 490–560-nm emission; Cy3-561-nm excitation, 563–593-nm emission;

© 2015 Nature America, Inc. All rights reserved.

Texas Red-594-nm excitation, 597–647-nm emission; and Cy5–633-nm excitation, 637–758-nm emission. Suggested microscope objectives are plan-Apochromat dry 20× NA 0.75, dry 40× NA 0.8 and water-immersion 63× NA 1.3.

Software installation Verify that Bio-Formats (http://loci.wisc.edu/software/bio-formats) plug-ins are available for Fiji/ImageJ. Download a free academic version of Canopy Python 2.7 in the home directory on the remote host, and follow the installation instructions (http://docs.enthought.com/canopy/quick-start/install_linux.html). Canopy Python 2.7 is easy to install, and it has all the required packages for our FISSEQ software. Install the latest version of the ggplot2 and data.table packages in RStudio.

PROCEDURE

FISSEQ library construction in cultured cells or tissue sections ● TIMING 2–3 d

▲ CRITICAL All reagents and washes are at room temperature unless indicated otherwise.

1| To construct libraries for FISSEQ, follow option A for cultured adherent cells or follow option B for tissue sections.

(A) Cultured adherent cells on a glass-bottom dish ● TIMING 30 min

(i) Fix the cells using 2 ml of 10% formalin in PBS for 15 min at 25 °C.
(ii) Wash the cells with 2 ml of PBS three times.
(iii) Add 2 ml of 0.25% (vol/vol) Triton X-100 in DEPC-PBS for 10 min, or 70% (vol/vol) ethanol for 2 min. Triton X-100 tends to maintain the subcellular structures better than 70% (vol/vol) ethanol.
(iv) Wash the cells with 2 ml of PBS three times.
(v) (Optional) Some cell types may require acid treatment for improved permeabilization: add 0.1 N HCl in DEPC-treated H_2O for 10 min, followed by three PBS washes (**Supplementary Fig. 2**).
? TROUBLESHOOTING

(B) Tissue sections on a glass-bottom dish ● TIMING 1 h

(i) Mount 10–20-μm-thick formalin-fixed tissue sections onto an RNase-free glass coverslip using a standard mounting procedure.
(ii) Remove the glass coverslip attached to a MatTek glass-bottom dish by gently pressing around the coverslip with a razor blade.
(iii) Attach the glass coverslip with a mounted tissue section to the MatTek dish using double-sided adhesive tape.
(iv) Wash the tissue section twice using DEPC-treated H_2O for 5 min each.
(v) Add 0.25% (vol/vol) Triton X-100 in DEPC-treated H_2O for 15 min, and aspirate.
(vi) Wash the sample with DEPC-treated H_2O twice.
(vii) Add 200 μl of 0.1% (wt/vol) pepsin in 0.1 N HCl for up to 10–30 min. Most tissue sections are permeabilized after 10–15 min. We recommend optimizing the permeabilization conditions for each tissue type.
(viii) Wash the tissue sections with 2 ml of PBS three times to inactivate pepsin.
? TROUBLESHOOTING

2| Prepare an RT mixture on ice, as indicated below, with and without reverse transcriptase.

▲ CRITICAL STEP Chilling the assembled mix to 4 °C before RT improves the efficiency of primer annealing.

Component	Amount (μl)	Final
DEPC-H_2O	159	
M-MuLV RT buffer, 10×	20	1×
dNTP, 25 mM	2	250 μM
Aminoallyl-dUTP, 4 mM	2	40 μM
RT primer, 100 μM (/5Phos/TCTCGGGAACGCTGAAGANNNNNN)	5	2.5 μM
RNase inhibitor (40 U μl^{1})	2	0.4 U μl^{-1}
M-MuLV reverse transcriptase (100 U μl^{-1})	10	5 U μl^{-1}
Total	200	

3| Incubate the specimen with the reaction mixture for 10 min at 4 °C, and then transfer it to 37 °C overnight. Typically, 1–2 h is sufficient, but more time may be required for thicker samples. Aspirate and wash the specimen with PBS once.

© 2015 Nature America, Inc. All rights reserved.

4| To cross-link cDNA molecules containing aminoallyl-dUTP, add 20 μl of reconstituted BS(PEG)9 in 980 μl of PBS to the sample for 1 h at room temperature.

5| Aspirate and wash the sample with PBS and quench it with 1 M Tris (pH 8.0) for 30 min.
■ **PAUSE POINT** The sample can be stored in PBS for up to 1 week at 4 °C.

6| Aspirate and add 10 μl of DNase-free RNase and 5 μl of RNase H in 1× RNase H buffer for 1 h at 37 °C (**Supplementary Fig. 3**).
▲ **CRITICAL STEP** Skipping this step results in few amplicons.

7| Rinse the sample with 2 ml of nuclease-free H_2O twice to remove traces of phosphate.

8| Prepare a CircLigase reaction mixture on ice, as tabulated below, and add it to the glass-bottom dish containing the sample.

Component	Amount (μl)	Final
Nuclease-free H_2O	128	
CircLigase buffer, 10×	20	1×
$MnCl_2$, 50 mM	10	2.5 mM
Betaine, 5 M	40	0.5 M
CircLigase II (100 U μl^{-1})	2	1 U μl^{-1}
Total	200	

9| Place the glass-bottom dish in a tightly sealed plastic container or in a Ziploc bag with moist wipes, and incubate it at 60 °C for 1 h. If a longer reaction time is desired, 1 ml of mineral oil can be layered on top of the sample.

10| Aspirate the reaction mixture, and wash with PBS. Mineral oil can be removed using PBS with 0.1% (vol/vol) Triton X-100.
■ **PAUSE POINT** The sample can be stored in PBS at 4 °C indefinitely.

11| Add 200 μl of RCA primer hybridization buffer containing 500 nM RCA primer to the glass-bottom dish and incubate at 60 °C for 1 h.

12| Aspirate and wash the sample with RCA hybridization buffer at 60 °C for 10 min.

13| Aspirate and wash the sample with 2× SSC, 1× SSC and PBS once each.

14| Prepare an RCA reaction mixture on ice, as tabulated below. Add this mixture to the sample and incubate it overnight at 30 °C. Additional dNTP (up to 10 μl) and ϕ29 DNA polymerase (up to 10 μl) can enhance the fluorescence signal from DNA amplicons.
▲ **CRITICAL STEP** Aminoallyl-dUTP is required for cross-linking and should not be omitted.

Component	Amount (μl)	Final
Nuclease-free H_2O	174	
ϕ29 buffer, 10×	20	1×
dNTP, 25 mM	2	250 μM
Aminoallyl-dUTP, 4 mM	2	40 μM
ϕ29 DNA polymerase (100 U μl^{-1})	2	1 U μl^{-1}
Total	200	

© 2015 Nature America, Inc. All rights reserved.

15| To cross-link cDNA molecules containing aminoallyl-dUTP, wash them gently with PBS, add 20 µl of reconstituted BS(PEG)9 in 980 µl of PBS to the sample and incubate the mixture for 1 h at room temperature.
▲ **CRITICAL STEP** BS(PEG)9 expires after 2–3 weeks with multiple freeze-thaw cycles, and using expired BS(PEG)9 can lead to unstable amplicons and poor sequencing results.

16| Wash the sample with PBS, aspirate and add 1 M Tris, pH 8.0, for 30 min.
■ **PAUSE POINT** Store the sample in PBS at 4 °C for up to 4 weeks.

17| Aspirate and add 2.5 µM control probe in 200 µl of 5× SASC, preheated to 80 °C, to the sample and incubate the mixture for 10 min at room temperature. Use the adapter- or rRNA–specific probes as positive controls to image all amplicons or rRNA amplicons, respectively. RT-negative controls should not produce any amplicons.

18| Wash the sample two times for 1 min each with 1 ml of 1× instrument buffer. If you are using adapter sequence–specific probe, proceed directly to Step 19 for imaging. If you are using the biotinylated rRNA probes, incubate in 2 µg ml^{-1} streptavidin–Alexa Fluor in PBS for 5 min, followed by three 2-ml PBS washes before continuing with Step 19.

19| Image on a microscope and inspect the amplicon density and distribution. Amplicons should be distributed uniformly throughout the sample across the glass-bottom dish. Obtain an axial view, and check to see whether the amplicon density is similar between regions near the glass and cell surface.
▲ **CRITICAL STEP** The sample can be imaged while immersed in 1× instrument buffer. If an alternative immersion liquid is used, do not add Tris-EDTA or other chelating agents.
? TROUBLESHOOTING

20| Aspirate and incubate the sample twice for 5 min each in 1 ml of strip buffer at room temperature, preheated to 80 °C.

21| Wash the sample twice for 5 min each with 1 ml of 1× instrument buffer at room temperature.
■ **PAUSE POINT** We have kept samples in 1× instrument buffer at 4 °C for up to several months without suffering a substantial loss in the fluorescence signal.

SOLiD sequencing-by-ligation ● TIMING 10 d for 30 cycles
22| Clamp the sample firmly to the microscope stage, and use cyanoacrylate adhesive to secure any potential sources of movement, such as adjustable stage inserts. Cyanoacrylate adhesive can be applied directly to metal components, and it can be removed with acetone after sequencing.
▲ **CRITICAL STEP** Use only optical-grade cyanoacrylate adhesive, as standard cyanoacrylate adhesives degas and ruin nearby objectives.

23| Add 2.5 µM sequencing primer N in 200 µl of 5× SASC, preheated to 80 °C, to the sample and incubate the mixture for 10 min at room temperature. Aspiration can be performed using a vacuum aspirator or a flexible plastic catheter attached to a syringe.

24| Wash the sample two times for 1 min each with 1 ml of 1× instrument buffer at room temperature.

25| Sequence the sample by adding a freshly prepared T4 DNA ligation mixture and incubating it for 45 min at room temperature.

Component	Amount (µl)	Final
Nuclease-free H_2O	165	
T4 DNA ligase buffer, 10×	20	1×
T4 DNA ligase, 120 U $µl^{-1}$	10	6 U $µl^{-1}$
SOLiD sequencing oligos		
(dark purple tube from the SOLiD ToP sequencing kit)	5	
Total	200	

© 2015 Nature America, Inc. All rights reserved.

26| Wash the sample four times for 5 min each with 1 ml of 1× instrument buffer at room temperature.

27| Acquire images.
▲ **CRITICAL STEP** The first ligation cycle for recessed primers *N*–2, *N*–3 and *N*–4 produces a fluorescence signal in just one channel. These images should not be included in the final data set.
? TROUBLESHOOTING

28| Aspirate and cleave the fluorophore by incubating the sample two times for 5 min each in cleave solution 1, and then two times for 5 min each in reconstituted cleave solution 2.1.
? TROUBLESHOOTING

29| Aspirate and wash the sample three times for 5 min each with 1 ml of 1× instrument buffer.
■ **PAUSE POINT** The sample is stable for 2–3 d in 1× instrument buffer at room temperature.

30| Repeat Steps 25–29 up to a total of seven cycles.

31| Incubate the sample four times for 5 min each in 1 ml of strip buffer, preheated to 80 °C.

32| Wash the sample two times for 1 min each with 1 ml of 1× instrument buffer.
■ **PAUSE POINT** The sample is stable for 2–3 d in 1× instrument buffer at room temperature.

33| Repeat Steps 23–32 using different sequencing primers (*N*–1, *N*–2, *N*–3 and *N*–4).
? TROUBLESHOOTING

Image pre-processing ● TIMING 6–12 h
34| If necessary, use ImageJ to crop image stacks for faster 3D deconvolution.

35| Determine the optimal 3D deconvolution parameters using a smaller cropped test image from the experiment. In Huygens Professional, we typically use a Nyquist sampling rate of 1.7, CMLE mode, 5–10 iterations and a signal-to-noise ratio of 2–5.

36| Deconvolve all sequencing images, and save images as .ics/.ids files with the following names in a folder named 'decon_images' (**Supplementary Fig. 4**). Filename: <Position>_<Primer #>_<Ligation #>_<Date__Time>.<ext>; Position: dinucleotide position as two-digit integers 01 to 30; Primer number: N followed by one-digit integers N0 to N4; Cycle number: ligation cycle per primer from 1 to 7; Date/time: An alphanumeric string using underscores; and File extension: .ics and .ids.
? TROUBLESHOOTING

Image analysis ● TIMING 6–12 h
▲ **CRITICAL** Some users of our method may have little or no background in bioinformatics. Here we introduce common computational environments and tools, but novice users should obtain additional help from experienced users, network administrators and online resources (i.e., http://www.ee.surrey.ac.uk/Teaching/Unix/).

37| Download fisseq.zip (http://arep.med.harvard.edu/FISSEQ_Nature_Protocols_2014/) and copy it to a remote host using a command-line terminal on a PC.

```
local:~$ scp fisseq.zip <user@remote_host_name:~/>
```

▲ **CRITICAL STEP** One must have an account to a designated remote host. Ask the network administrator at your institution.

38| Download and unzip decon_images.zip (http://arep.med.harvard.edu/FISSEQ_Nature_Protocols_2014/). Copy the decon_images folder (Step 36) to a scratch space on the remote host.

```
local:~$ scp -r ~/decon_images/ <user@remote_host_name:scratch_space/>
```

▲ **CRITICAL STEP** Analysis of multiple high-resolution image stacks requires a large amount of disk space. Contact your network administrator for the location of a temporary scratch space.

© 2015 Nature America, Inc. All rights reserved.

39| Log on to the remote host and submit a job request to work on a high-memory queue interactively. We recommend at least 100 GB (memory below is in MB).

```
local:~$ ssh <user@remote_host_name>

remote:~$ bsub -R "rusage[mem=100000]" -q <queue_name> -Is bash
```

▲ **CRITICAL STEP** Running CPU or memory-intensive tasks incorrectly can bring down the remote host. Make sure that you are working on a designated node. Contact your network administrator for more information before proceeding.

40| Unzip fisseq.zip and change the working directory to fisseq.

```
remote:~$ unzip fisseq.zip

remote:~$ cd fisseq
```

▲ **CRITICAL STEP** Working from folders other than ~/fisseq results in missing file errors when entering our commands as written below.

41| Download and decompress the RefSeq-to-Gene ID conversion table.

```
remote:~/fisseq$ wget ftp://ftp.ncbi.nih.gov/gene/DATA/gene2refseq.gz

remote:~/fisseq$ gzip -d gene2refseq.gz
```

42| Download the organism-specific RefSeq RNA FASTA file and unzip the file.

```
remote:~/fisseq$ wget

ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/mRNA_Prot/human.rna.fna.gz

remote:~/fisseq$ gzip -d human.rna.fna.gz

## Use the following address for mouse or rat:

## ftp.ncbi.nlm.nih.gov/refseq/M_musculus/mRNA_Prot/mouse.rna.fna.gz

## ftp.ncbi.nlm.nih.gov/refseq/R_norvegicus/mRNA_Prot/rat.rna.fna.gz
```

43| Build the reference index of [ref_name] in color space. Here [ref_name] is refseq_human. This process can take several hours.

```
remote:~/fisseq$ bowtie-build -C -f human.rna.fna refseq_human
```

? TROUBLESHOOTING

44| Start MATLAB and add a search path:

```
remote:~/fisseq$ matlab

>> addpath('~/fisseq', '~/fisseq/bfmatlab')
```

45| Define the input and output directories and run image registration (**Supplementary Fig. 5**).

```
>> input_dir='<scratch_space>/decon_images/'

>> output_dir='registered_images/'

>> register_FISSEQ_images(input_dir,output_dir,10,0.1,1)

>> quit()
```

© 2015 Nature America, Inc. All rights reserved.

Set the mumber of blocks per axis for local registration (default = 10); set the fraction overlap between neighboring blocks (default = 0.1); and adjust the alignment precision, where 10 will register images to 1/10 of a pixel (default = 1).
? TROUBLESHOOTING

46| Copy files in ~/fisseq/registered_images/ to a PC. Use ImageJ to open TIFF files (File > Import > Bio-Formats) as a time series, and check alignment in channel 4 by scrolling through the timeline (**Supplementary Videos 1–4**). Maximum-projected TIFF files (channel 4 is a composite of channels 0–3); Routput.mat file: block-wise registration offsets between bases; Rchadj.mat file: block-wise chromatics shifts as a matrix; and Rtadj.mat file: registration offsets over time for the whole image (not block-wise).
? TROUBLESHOOTING

47| Start python, and write base calls to read_data_*.csfasta. The maximum number of missing base calls allowed per read is 6 by default. * denotes an automatically generated time stamp.

```
remote:~/fisseq$ python

>>> import FISSEQ

>>> FISSEQ.ImageData('registered_images','.',6)

>>> quit()
```
? TROUBLESHOOTING

48| Align reads to refseq_human (Step 43) using Bowtie 1.0 or earlier, and write mapped reads to bowtie_output.txt. The exact name of read_data_*.csfasta can be determined by listing files in the directory (ls -l).

```
remote:~/fisseq$ bowtie -C -n 3 -l 15 -e 240 -a -p 12 -m 20 --chunkmbs 200 -f --best
--strata --refidx refseq_human read_data_*.csfasta bowtie_output.txt
```
? TROUBLESHOOTING

49| Spatially cluster the Bowtie reads (Step 48), annotate clusters using gene2refseq (Step 41) and write to results.tsv. The default kernel size of 3 performs a 3 × 3 dilation before clustering.

```
remote:~/fisseq$ python

>>> import FISSEQ

>>> G = FISSEQ.ImageData('registered_images',None,6)

>>> FISSEQ.AlignmentData('bowtie_output.txt',3,G,'results.tsv',

   'human.rna.fna','gene2refseq','9606')

>>> quit()

## Use the following command for mouse or rat:

>>> FISSEQ.AlignmentData('bowtie_output.txt',3,G,'results.tsv',

   'mouse.rna.fna','gene2refseq','10090')

>>> FISSEQ.AlignmentData('bowtie_output.txt',3,G,'results.tsv',

   'rat.rna.fna','gene2refseq','10116')
```

Data analysis ● TIMING 1 d

▲ **CRITICAL** Data analysis can be done on any software package, but R is convenient for interactive analysis and high-quality graphs[23]. Novice users may find RStudio more intuitive than the command-line interface. We provide a sample R session containing a sample data set and a list of commands (http://arep.med.harvard.edu/FISSEQ_Nature_Protocols_2014/).

© 2015 Nature America, Inc. All rights reserved.

50| Open the FISSEQ RStudio project file (*Menu* → *File* → *Open project*…).

51| Find the HISTORY tab on the upper right console window, and double-click on individual commands in order to re-execute the previous R session (**Supplementary Fig. 6**) and learn how to: import and filter data using a specific criterion (i.e., cluster size); plot a distribution of reads by a specific criterion (i.e., RNA classes and strands); convert a table of reads into a table of gene expression level; correlate gene expression from different images; and find statistically enriched genes in different regions.
? TROUBLESHOOTING

? TROUBLESHOOTING
Troubleshooting advice can be found in **Table 3**.

TABLE 3 | Troubleshooting table.

Step	Problem	Possible reason	Solution
1	Cells wash away during PBS washes or fixation	Cell attachment dependent on Ca^{++}	Use PBS with Ca^{++}, or add formalin directly to the growth medium
	Membrane blebs	Diluting formalin using 1× PBS makes it hypotonic	Use 10× PBS and water to dilute the formalin
	Tissue sections falling apart or off slide	Pepsin overdigestion and/or small contact area between sample and glass	Shorter pepsin digestion; embed in Matrigel and re-fix
19	Few amplicons limited to the cell surface	Poor cell permeabilization	Use 0.1 N HCl after Triton or ethanol-based cell permeabilization
	Amplicons in no-RT control	RT primer is too long	Use shorter RT primers
	High background	Excess RCA primer	More stringent washes at Step 13
	Dim amplicons	Low template copy number per amplicon	More dNTP and ϕ29 enzyme at Step 14
19,27	Dim, fuzzy or stretched amplicons	Poor cross-linking	Fresh BS(PEG)9 at Step 15
28	White precipitate buildup	Silver reacts with chloride	Eliminate chloride-containing buffers
33	Progressive loss of signal	Photodamage to amplicons	Low laser exposure
36	Deconvolution takes too much time	Large images	Crop unused areas Smaller images Fewer iterations
43	Bowtie command is not found	Bowtie v1.0 environment is not set up	Check Bowtie version (which bowtie); ask administrator for assistance
45	MATLAB out of memory error	Low RAM Low heap space for Java virtual memory (VM)	Allocate >100 GB RAM Increase Java VM in java.opts
	Input or output folders are not found	Incorrect slash use with folder name	No slash before and one slash after
46	ImageJ does not open TIFF files correctly	Image dimensions are not correctly read	Check 'Group files…', 'Swap dim…' and 'Concatenate…' when importing
47	Cannot find input images	Undefined path	Registered image directory must be in ~/fisseq
	Extension error messages	Missing package	Use Canopy Python 2.7

(continued)

© 2015 Nature America, Inc. All rights reserved.

TABLE 3 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
48	Bowtie is not found	Bowtie 1.0 is not loaded	Check available versions and load Bowtie 1.0 or earlier versions
	Extra parameter(s) error	Typo, or option flags are not in the correct order	Copy and paste the command from Step 48
51	Unexpectedly high number of anti-sense mRNA reads	Noisy image, many missing or incorrect base calls at the 3′ end	Obtain better images, use deconvolution to reduce noise, sequence longer reads, trim reads or increase the cluster size threshold

● TIMING

Steps 1–21, FISSEQ library construction: 2–3 d
Steps 22–33, sequencing and imaging: 10 d
Steps 34–36, image pre-processing: 6–12 h
Steps 37–49, image analysis: 6–12 h
Steps 50 and 51, data analysis: 1 d

ANTICIPATED RESULTS

The size of subcellular cDNA amplicons is slightly larger than the diffraction limit after 3D deconvolution. At 20× NA 0.75, the diameter of cDNA amplicons is ~400–800 nm after image deconvolution. A typical amplicon contains hundreds of fluorescent probe–binding sites, and this results in images that are 20–50 times brighter and that have a markedly improved signal-to-noise ratio than single-molecule FISH. A good FISSEQ library should yield many intensely bright amplicons that are distinct from cell debris and spurious amplification products. If long exposure time and high gain have to be used to visualize objects, it is likely that they represent contamination, reaction precipitates or cell debris.

When fluorescent probes are stripped, nearly all of the fluorescence is completely removed, except possibly in the nucleus. Stripping is a good way to distinguish a DNA amplicon from fluorescent debris, and we recommend alternately hybridizing the sample with FAM, Cy3 or Cy5 probes while the sample is still on the microscope. If the fluorescent object is a DNA amplicon, it should fluoresce in distinct colors sequentially with little or no cross-talk. The amplicon density varies depending on the cell size, but we typically see several hundreds of amplicons per cell in cultured cell lines (i.e., iPSCs, fibroblasts, HeLa cells and bipolar neurons). We have detected up to 4,000 amplicons using synthetic DNA per cell in fibroblasts, suggesting that the RT efficiency may be a limiting factor.

The signal-to-noise ratio from SOLiD sequencing-by-ligation is high, especially for early ligation cycles. The quality drops after the fourth re-ligation cycle for each primer, and the image quality degrades significantly after 25 total cycles. Much of the image degradation results from the laser-induced damage during imaging. Typically, unimaged regions remain pristine even after 30 cycles of sequencing, and it may be possible to obtain a longer read length with appropriate free-radical scavengers in the imaging buffer, but we have not attempted this yet.

Depending on the camera sensor size, density and bit depth, one image stack containing multiple optical planes across four channels can be 800 MB–2 GB per field of view. Our image registration software then creates a separate folder containing TIFF images (five channels per base) of 20–50 M in size. Once our software processes and analyzes the images, it generates a tab-delimited file containing the gene ID, name, cluster size, strand, class, base quality, alignment quality, color space sequence and *x*-*y* position. We recommend performing a quick data check by selecting a gene cluster size of >5 to compare the number of sense and anti-sense reads, and we also recommend comparing the number of reads from different RNA classes. Typically, >90% of all reads should map to the positive sense strand. The rRNA read should comprise 50–80% of the total number of reads. We typically get 15,000–40,000 reads per image containing 30–50 cells. Regional or subcellular localization is measured in statistically significant enrichment scores, rather than absolute counts, owing to a small number of reads distributed over a large area. We recommend making B&W image masks on the basis of the cell morphology, DAPI stains, immunohistochemistry and other types of spatial masks, and measuring the relative enrichment of individual genes using Fisher's exact test or other similar tests[23]. With a high read density, it may be possible to use unsupervised local clustering of reads for regional identification of biological processes[2].

© 2015 Nature America, Inc. All rights reserved.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS This study was funded by US National Institutes of Health (NIH) Centers of Excellence in Genomic Sciences (CEGS) grant no. P50 HG005550. J.H.L. and co-workers were funded by National Heart, Blood and Lung Institute (NHBLI) grant no. RC2HL102815, by the Allen Institute for Brain Science and by National Institute of Mental Health (NIMH) grant no. MH098977. E.R.D. was funded by NIH grant no. GM080177 and by National Science Foundation (NSF) Graduate Research Fellowship grant no. DGE1144152.

AUTHOR CONTRIBUTIONS J.H.L. and E.R.D. conceived FISSEQ library construction, sequencing, image analysis and bioinformatics. J.S., R.K., J.L.Y., B.M.T., H.S.L. and J.A. provided key feedbacks during the FISSEQ method development. R.T. and T.C.F. assisted with automated microscopy and image analysis. K.Z. and G.M.C. oversaw the project. J.H.L. wrote the paper, and E.R.D. wrote the FISSEQ software.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

© 2015 Nature America, Inc. All rights reserved.

1. Rifai, N., Gillette, M.A. & Carr, S.A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24**, 971–983 (2006).
2. Jaitin, D.A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
3. Battich, N., Stoeger, T. & Pelkmans, L. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat. Methods* **10**, 1127–1133 (2013).
4. Lein, E.S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
5. Zeng, H. *et al.* Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149**, 483–496 (2012).
6. Diez-Roux, G. *et al.* A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9**, e1000582 (2011).
7. Femino, A.M., Fay, F.S., Fogarty, K. & Singer, R.H. Visualization of single RNA transcripts *in situ*. *Science* **280**, 585–590 (1998).
8. 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).
9. Levsky, J.M., Shenoy, S.M., Pezo, R.C. & Singer, R.H. Single-cell gene expression profiling. *Science* **297**, 836–840 (2002).
10. 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).
11. Choi, H.M. *et al.* Programmable *in situ* amplification for multiplexed imaging of mRNA expression. *Nat. Biotechnol.* **28**, 1208–1212 (2011).
12. Ke, R. *et al.* *In situ* sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
13. Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. *In situ* detection and genotyping of individual mRNA molecules. *Nat. Methods* **7**, 395–397 (2010).
14. Larsson, C. *et al.* *In situ* genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nat. Methods* **1**, 227–232 (2004).
15. Lagunavicius, A. *et al.* Novel application of Phi29 DNA polymerase: RNA detection and analysis *in vitro* and *in situ* by target RNA-primed RCA. *RNA* **15**, 765–771 (2009).
16. Merkiene, E., Gaidamaviciute, E., Riauba, L., Janulaitis, A. & Lagunavicius, A. Direct detection of RNA *in vitro* and *in situ* by target-primed RCA: the impact of *E. coli* RNase III on the detection efficiency of RNA sequences distanced far from the 3′-end. *RNA* **16**, 1508–1515 (2010).
17. Lee, J.H. *et al.* A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet.* **5**, e1000718 (2009).
18. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
19. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
20. Grun, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11**, 637–640 (2014).
21. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
22. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).
23. Lee, J.H. *et al.* Highly multiplexed subcellular RNA sequencing *in situ*. *Science* **343**, 1360–1363 (2014).
24. Adiconis, X. *et al.* Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* **10**, 623–629 (2013).
25. Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
26. Frumkin, D. *et al.* Amplification of multiple genomic loci from single cells isolated by laser micro-dissection of tissues. *BMC Biotechnol.* **8**, 17 (2008).
27. Lovatt, D. *et al.* Transcriptome *in vivo* analysis (TIVA) of spatially defined single cells in live tissue. *Nat. Methods* **11**, 190–196 (2014).
28. Schmid, M.W. *et al.* A powerful method for transcriptional profiling of specific cell types in eukaryotes: laser-assisted microdissection and RNA sequencing. *PLoS ONE* **7**, e29685 (2012).
29. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).
30. Ramskold, D. *et al.* Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
31. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
32. Avital, G., Hashimshony, T. & Yanai, I. Seeing is believing: new methods for *in situ* single-cell transcriptomics. *Genome Biol.* **15**, 110 (2014).
33. Buxbaum, A.R., Wu, B. & Singer, R.H. Single β-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* **343**, 419–422 (2014).
34. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).
35. Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
36. Itzkovitz, S., Blat, I.C., Jacks, T., Clevers, H. & van Oudenaarden, A. Optimality in the development of intestinal crypts. *Cell* **148**, 608–619 (2012).
37. Hansen, C.H. & van Oudenaarden, A. Allele-specific detection of single mRNA molecules *in situ*. *Nat. Methods* **10**, 869–871 (2013).
38. Porreca, G.J. *et al.* Multiplex amplification of large sets of human exons. *Nat. Methods* **4**, 931–936 (2007).
39. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **110**, 14024–14029 (2013).
40. Li, J.B. *et al.* Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
41. Zhang, K. *et al.* Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat. Methods* **6**, 613–618 (2009).
42. Michael, W.M., Choi, M. & Dreyfuss, G. A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* **83**, 415–422 (1995).
43. Kaposi-Novak, P., Lee, J.S., Mikaelyan, A., Patel, V. & Thorgeirsson, S.S. Oligonucleotide microarray analysis of aminoallyl-labeled cDNA targets from linear RNA amplification. *Biotechniques* **37**, 580, 582–586, 588 (2004).
44. Nanda, J.S. & Lorsch, J.R. Labeling a protein with fluorophores using NHS ester derivitization. *Methods Enzymol.* **536**, 87–94 (2014).
45. Mardis, E.R. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* **9**, 387–402 (2008).
46. Massingham, T. & Goldman, N. Error-correcting properties of the SOLiD Exact Call Chemistry. *BMC Bioinformatics* **13**, 145 (2012).
47. Applied Biosystems. *SOLiD System Accuracy with the Exact Call Chemistry Module* (https://www3.appliedbiosystems.com/cms/groups/global_marketing_group/documents/generaldocuments/cms_091372.pdf).
48. Itzkovitz, S. & van Oudenaarden, A. Validating transcripts with probes and imaging technology. *Nat. Methods* **8**, S12–S19 (2011).
49. Eliceiri, K.W. *et al.* Biological imaging software tools. *Nat. Methods* **9**, 697–710 (2012).
50. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
51. Kankaanpaa, P. *et al.* BioImageXD: an open, general-purpose and high-throughput image-processing platform. *Nat. Methods* **9**, 683–689 (2012).
52. Pawley, J.B. *Handbook of Biological Confocal Microscopy* 3rd edn. (Springer, 2006).