Validating transcripts with probes and imaging technology

Shalev Itzkovitz[1,2] & Alexander van Oudenaarden[1–3]

High-throughput gene expression screens provide a quantitative picture of the average expression signature of biological samples. However, the analysis of spatial gene expression patterns with single-cell resolution requires quantitative *in situ* measurement techniques. Here we describe recent technological advances in RNA fluorescence *in situ* hybridization (FISH) techniques that facilitate detection of individual fluorescently labeled mRNA molecules of practically any endogenous gene. These methods, which are based on advances in probe design, imaging technology and image processing, enable the absolute measurement of transcript abundance in individual cells with single-molecule resolution.

© 2011 Nature America, Inc. All rights reserved.

High-throughput measurements of gene expression on a genomic scale using microarray technology or high-throughput sequencing contributed tremendously to our understanding of how genetic networks coordinately function in normal cells and tissues and how they malfunction in disease. Such measurements allow one to infer the function of genes based on their expression patterns[1], to detect which genes have altered expression in disease[2] and to identify expression signatures that are predictive of cancer progression[3,4]. However, the variability in single-cell gene expression in most biological systems and especially in tissues and tumors suggests that bulk transcriptome measurements should be complemented by techniques aimed at characterizing gene expression programs in individual cells[5]. In this Review we will describe advances in single-molecule transcript imaging, which yield integer counts of transcripts in single cells in suspension and in intact tissues.

Biological samples are inherently heterogeneous

Bulk transcriptome measurements inform on the average gene expression in a sample. Thus, in a heterogeneous sample, containing several cell types with different gene expression signatures, only the most abundant signature will be captured. Such heterogeneity is present in practically any biological sample. Bacterial and yeast cells that are derived from isogenic monoclonal populations have been shown to have pronounced cell-to-cell variation in the expression of many genes, stemming from stochastic events such as bursting transcriptional dynamics and cell-cycle dependence[6]. Expression heterogeneity is even more pronounced in tissues, which are usually composed of several types of cells with profoundly different gene expression programs. In many epithelial tissues, such as the skin and the intestine, there is a clear hierarchical partition into stem cells and diverse differentiated epithelial progenies, each of which has distinct phenotypic and morphological features. The precise location of cells in tissue translates to constant changes in the levels of niche-secreted morphogens, which give rise to position-modulated gene expression programs. Thus, two adjacent cells could have dramatically different expression programs.

Solid tumors represent a particular case of cell heterogeneity. Most solid tumors consist of a mixture of cancer and stromal cells. Additionally, cancer cells are often profoundly diverse not only in their transcript content but also in their genotype. This diversity stems from increased mutation rates, rapid cell proliferation and spatially varying selection forces. Single cells from a wide range of colorectal cancer cell lines change their chromosomal copy number on average

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Correspondence should be addressed to A.v.O. (avano@mit.edu).

PUBLISHED ONLINE 30 MARCH 2011; DOI:10.1038/NMETH.1573

once every five cell divisions *in vitro*[7], and there is a dramatic heterogeneity in copy numbers observed in tissue cross-sections[8,9]. A bulk measurement of gene expression in tumors only captures the most dominant tumor clone and is masked by stromal signals, therefore providing only partial information on the different expression signatures coexisting in the tumor. Moreover, some tumors have been shown to contain a minority of cells with increased oncogenic ability—termed cancer stem cells[10]. Bulk measurements of tumors cannot capture the expression signature of such minor populations.

Attempts to address this underlying heterogeneity are based on enriching for certain subpopulations from tissues or tumors. An example of such enrichment is the use of laser capture microdissection, which enables extraction of regions in a tissue based on morphological features or gene expression markers[11]. This technique is frequently applied to selectively collect tumor cells rather than a mixture of tumor and stroma cells for downstream expression analysis. Alternatively, a fluorescence-activated cell sorter can be used to specifically isolate cells expressing a small number of defined gene expression markers. These cells can then be used for either bulk measurements or for single-cell measurements that require cell lysis, such as quantitative reverse transcription (RT)-PCR[12] or digital RT-PCR[13]. Similarly, sorting GFP-positive cells from transgenic mice expressing GFP under the control of a cell type–specific promoter of interest enables enrichment for the cells expressing this gene and allows characterization of the expression program of these particular cells, as has been done with tissue stem cells[14]. The main drawback of these methods is the fact that they involve dissociating the tissue and thus result in the complete loss of spatial information. Moreover, specific expression in only one cell type is uncommon, and thus the enriched sample would generally still consist of a mixture of different cell populations. In addition, the variability resulting from the reverse transcription and the exponential PCR amplification steps often limits the sensitivity and reproducibility of these approaches[12]. Thus quantitatively unraveling the gene expression programs of single cells in heterogeneous samples while preserving spatial context requires methods for transcript measurement in intact tissues.

Traditional approaches for measuring *in situ* gene expression

The traditional method for measuring gene expression in intact tissues is RNA *in situ* hybridization (ISH) and its variant RNA fluorescence *in situ* hybridization (FISH)[15]. FISH methods have originally been developed for DNA analysis[16] and are based on the specific binding of long fluorescently labeled oligonucleotide probes to their complementary sequences in fixed and permeabilized samples. Although there are many nuances in FISH techniques, most entail the following steps (**Fig. 1**). Tissues or cells are fixed, permeabilized and then hybridized with fluorescently labeled oligonucleotide probes. Probes are either directly coupled to fluorescent molecules or coupled to haptens, typically biotin or digoxigenin, to which either avidin or antibody to digoxigenin bind, respectively. These are, in turn, either directly conjugated to fluorophores or to enzymes such as alkaline phosphatase or horseradish peroxidase, which generate either fluorescent (in FISH) or chromogenic (in ISH) products from specific added substrates. Indirect labeling offers the advantage that signal can be amplified by using primary and secondary antibodies, of which only the secondary antibody elicits a light-emitting reaction[17].

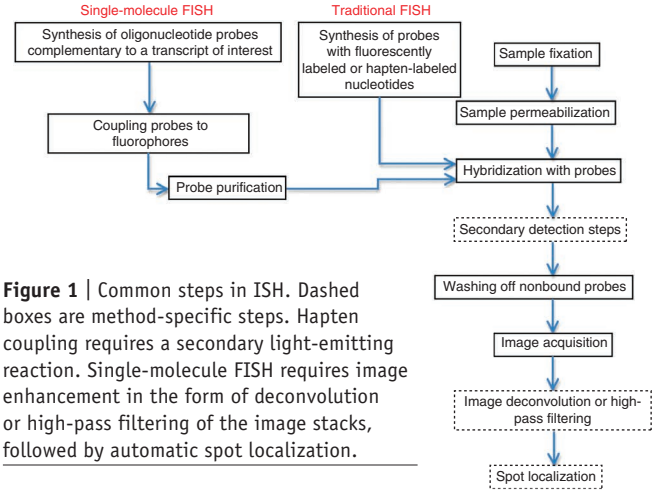

Figure 1 | Common steps in ISH. Dashed boxes are method-specific steps. Hapten coupling requires a secondary light-emitting reaction. Single-molecule FISH requires image enhancement in the form of deconvolution or high-pass filtering of the image stacks, followed by automatic spot localization.

Traditionally, fluorophores have been coupled to the oligonucleotide probes by enzymatic means such as nick translation or *in vitro* transcription, methods in which dye-modified nucleotides are stochastically incorporated along the probe (**Fig. 2a**). Although detection of single mRNA molecules using traditional FISH has been reported for a specific probe in *Drosophila melanogaster*[18], in most cases traditional ISH and FISH techniques can only give qualitative information on gene expression. One drawback is the random distribution of dyes along the linear sequence, which has been shown to impede the hybridization quality by destabilizing the probe-target duplex, often leading to mutual quenching of adjacent dyes[19]. Also, long probes are poorly permeant, and have high background and low sensitivity. As it has been estimated that more than 80% of yeast genes are expressed at lower than two mRNA copies per cell[20], it would be impossible to study the majority of the eukaryotic transcriptome with traditional FISH.

Single-molecule transcript imaging

Recent improvements in probe designs, imaging technology and image processing software have enabled highly specific and robust hybridization of fluorescently labeled probes to an arbitrary transcript of interest with high spatial localization, yielding diffraction-limited spots that are visible under a fluorescence microscope (**Table 1**). Thus, an integer number of transcripts in individual cells can be determined by simply counting fluorescent dots. Moreover, the precise subcellular localization of the individual transcripts can be detected.

Short probes labeled with multiple fluorophores. Robert Singer and colleagues pioneered single mRNA molecule imaging techniques[21]. Their key improvement was the replacement of long probes with several 50-base-pair (bp) probes that are complementary to sequential parts of the target mRNA and are each coupled to typically 3–5 fluorescent dyes at predefined positions (**Fig. 2b**). A minimal preset spacing between incorporated dyes along the linear oligonucleotide sequence prevented quenching of adjacent dyes[22]. Coupled probes were separated from noncoupled probes and free dyes using chromatography. Singer and colleagues optimized the G+C content of all probes to be as close as possible to 50% and

© 2011 Nature America, Inc. All rights reserved.

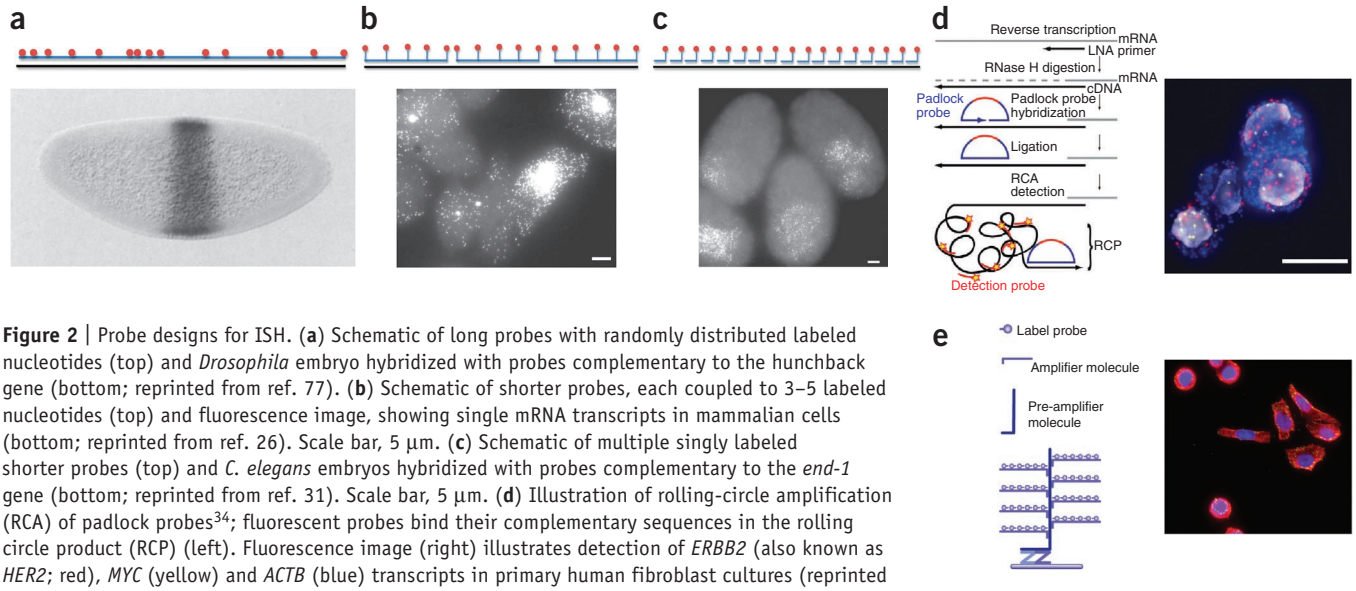

Figure 2 | Probe designs for ISH. (**a**) Schematic of long probes with randomly distributed labeled nucleotides (top) and *Drosophila* embryo hybridized with probes complementary to the hunchback gene (bottom; reprinted from ref. 77). (**b**) Schematic of shorter probes, each coupled to 3–5 labeled nucleotides (top) and fluorescence image, showing single mRNA transcripts in mammalian cells (bottom; reprinted from ref. 26). Scale bar, 5 µm. (**c**) Schematic of multiple singly labeled shorter probes (top) and *C. elegans* embryos hybridized with probes complementary to the *end-1* gene (bottom; reprinted from ref. 31). Scale bar, 5 µm. (**d**) Illustration of rolling-circle amplification (RCA) of padlock probes[34]; fluorescent probes bind their complementary sequences in the rolling circle product (RCP) (left). Fluorescence image (right) illustrates detection of *ERBB2* (also known as *HER2*; red), *MYC* (yellow) and *ACTB* (blue) transcripts in primary human fibroblast cultures (reprinted from ref. 34). Scale bar, 20 µm. (**e**) Schematic of a branched DNA detection method using four probe sets. Multiple label probes bind multiple amplifier molecules that are in turn bound to a pre-amplifier molecule that binds the transcript-specific probe pairs. Image of *ERBB2* (green) and 18S rRNA (red) in SKBR3 cells (right; reprinted from http://www.panomics.com/).

imaged the hybridized samples with a fluorescence microscope that captures stacks of images every fraction of micrometers, using a high-numerical-aperture objective and a low-noise charge-coupled device (CCD) camera. These improvements yielded three-dimensional digital images in which the specific accumulation of fluorescent molecules in the small volume occupied by a single mRNA molecule appeared as diffraction-limited spots.

Several research groups have since used multiply labeled probes for single-mRNA counting to characterize transcript distribution in yeast[23–25] and in mammalian cells[26]. In mammalian tissue, this method has not yet allowed single-transcript resolution but has enabled the detection of transcription sites in paraffin-embedded human tumors[27]. The approach has some drawbacks, however: mainly a high variability in the number of probes bound to the target[21]. Although ideally each spot would be composed of the same number of probes, in fact most fluorescent spots have been estimated to originate from only one or two probes. This made it difficult to differentiate the true specific binding to the legitimate target from nonspecific binding. Because an inefficient binding of one or two probes can lead to a high variability in spot intensity when only a handful of probes are used, it is important to carefully select the probes using this method. An additional source of variability using probes designed to be coupled to several fluorescent molecules is in the number of fluorescent molecules actually coupled to each probe; it is difficult to separate fully coupled probes from partially coupled ones.

Short probes labeled with single fluorophores. Raj *et al.*[28] modified Singer's method to create probe libraries consisting of many short, 17–22-bp oligonucleotides (typically 48–96 oligos) labeled with only a single fluorophore at their 3′ termini[28] (**Fig. 2c**). This allowed more efficient purification, and the variability in spot intensity caused by inefficient binding of any single probe was lower when multiple probes were used. Optimization of probe design to have a uniform G+C content of around 45% and a minimum gap of three nucleotides between successive probes enabled robust specific hybridization with 48 probes or less. Raj *et al.*[28] demonstrated simultaneous transcript counting of three transcripts coupled to distinct fluorophores. The use of a specialized mounting medium containing an oxygen-scavenging system inhibited oxygen-dependent, light-initiated pathways that destroy fluorophores, thus increasing dye photostability. This is particularly important in the optical setup in which the same field of view is illuminated multiple times as the optical sections are gathered[22].

Several research groups have applied singly labeled probes to a wide range of samples, ranging from yeast[29] and mammalian cells[28,30] to *Drosophila*[28] and *Caenorhabditis elegans* embryos[31]. The approach is appealing mainly for its simplicity and generality, aided by a simple web interface to design optimal probes for arbitrary sequences (http://www.singlemoleculefish.com/). Probe libraries are typically prepared with a 96-position DNA synthesizer, pooled and then simultaneously coupled to a fluorophore of choice. This format facilitates additional flexibility by enabling coupling of the same library to several different fluorescent dyes for simultaneous hybridization to other genes of interest.

A current limitation is the difficulty in detecting short transcripts. Sunney Xie and co-workers recently demonstrated single-mRNA detection with only one 20-bp fluorescently labeled probe in *Escherichia coli*[32]. The ability to detect single transcripts in mammalian cells using one probe is much more challenging because of the larger volumes and increased off-target sequences involved. *In situ* detection of short transcripts using oligonucleotide probes can be achieved by using modified nucleic acids to increase specificity and by applying signal amplification to increase sensitivity. Next we will discuss these two approaches.

Short probes with modified backbones. When the target sequences are too short, one runs into specificity problems, as any single probe has a nonnegligible probability of binding a different target. Increasing the sensitivity and specificity of individual probes could enable the detection of very short transcripts such

© 2011 Nature America, Inc. All rights reserved.

Table 1 | Technologies for single-molecule transcript imaging

Method	Single-molecule resolution	Live-cell imaging	Advantages	Disadvantages	References
Traditional FISH	No	No	Probes can be designed for a wide range of genes	Limited sensitivity and specificity; impossible to detect very short transcripts	15
Probes labeled with multiple fluorophores	Yes	No	Generic approach suited for most transcripts; established protocols exist	High variability in the number of probes bound to target	21–26,39
Multiple probes labeled with single fluorophores	Yes	No	Generic approach suited for most transcripts; low variability in the number of probes bound to target; commercially available labeled probes	Limited in the detection of very short transcripts	28,30,31; http://info.biosearchtech.com/singlemoleculefish/
Rolling-circle amplification of padlock probes	Yes	No	Can detect very short probes and discern single-nucleotide differences	Protocol requires reverse transcription and amplification	34
Branched DNA probes	Yes	No	Commercially available labeled probes	Protocol requires amplification	38; http://www.panomics.com/
Quantum dot–labeled probes	Yes	Yes	Bright and photostable; huge spectral range	Large physical size is detrimental to target binding and can cause cell penetration problems	70–75
Subdiffraction microscopy	Yes	Yes	Achieves a spatial resolution of 20 nm	Slow; expensive instrumentation	64–68
MS2-GFP	Yes	Yes	No need for external interventions (for example, microinjection); yields spatial information	Requires generating transgenes; mRNAs tend to form clumps	43,44,78
Molecular beacons	Yes	Yes	High specificity with no clumping; labeled probes are commercially available	Requires microinjecting probes	52–54

© 2011 Nature America, Inc. All rights reserved.

as single microRNA molecules, which typically span 19–24 bp[33] as well as allow distinguishing between transcript variants that differ by single nucleotides[34]. This can be achieved with probes containing modified nucleic acids such as peptide nucleic acids (PNA) and locked nucleic acids (LNA). PNAs have an uncharged peptide-like backbone[35] and therefore hybridize more stably to RNA compared to DNA probes. LNA is a 2′-O, 4′-C-methylene-linked ribonucleotide derivative of RNA, enabling more specific hybridization with RNA and DNA, compared to DNA or RNA probes. PNA probes have been used to detect telomeres (repetitive hexamer sequences at chromosome ends)[36], whereas LNA probes facilitated the detection of microRNAs[33,37] but with the use of signal amplification.

Signal amplification of single-molecule probes. Detection of short transcripts is still hampered by insufficient fluorescence from a single bound probe. Amplifying the fluorescent signal can solve this sensitivity limitation. Single microRNA molecules have been detected *in situ* using a single LNA probe and an alkaline phosphatase–based signal amplification[33]. Larsson *et al.*[34] introduced the use of padlock probes that can distinguish transcripts that differ by only a single base pair. The authors first reverse-transcribed the mRNA into cDNA using LNA primers, then hybridized linear padlock probes to juxtaposed segments of the target sequence, enzymatically ligated them and used them as templates for rolling circle amplification by Phi29 DNA polymerase (**Fig. 2d**)[34]. This created a single strand of DNA containing tandem repeats of the padlock probe sequence that, after hybridization to fluorescently labeled probes, yielded bright diffraction-limited spots. Larsson *et al.*[34] demonstrated multiplex detection of up to three genes in cells and frozen mouse embryonic tissue.

In the branched DNA approach[38] hybridization is performed with four distinct probe sets: a gene-specific probe set composed of ten or more oligonucleotide probe pairs that are complementary to the target sequence, a pre-amplifier probe that hybridizes to gene-specific probes, multiple amplifier probes that hybridize to the pre-amplifier probe and labeled probes that attach to the amplifier probes. The resulting construct yields bright concentrated fluorescence (**Fig. 2e**). This technology is commercially named QuantiGene ViewRNA[38].

Spectral barcoding. The number of spectrally resolvable fluorophores, typically three, limits the number of simultaneously measured genes in single-transcript imaging techniques. 'Spectral barcoding' is an approach developed by Singer and colleagues to increase the number of simultaneously detected transcripts[21,39]. The technology, which was based on a similar method used for DNA FISH[40,41], entailed the division of the probe set that is complementary to a given transcript of interest into groups, each of which is coupled to a different fluorophore. By precisely registering the images from the different fluorescence channels, one can determine not only whether a spot is present or not, but also how many colors comprise it. With n spectrally resolvable fluorophores one can, in principle, achieve $2^n - 1$ different probe color combinations. This limit could be increased if the detection method is sensitive enough to detect not only whether a given fluorophore is present or absent at a diffraction-limited spot but also estimate the number of oligonucleotides coupled to the fluorophore of interest[39].

Single-molecule transcript imaging in live cells. A limitation of transcript imaging using FISH approaches is that these require the

samples to be fixed. The ability to temporally measure gene expression in living cells offers a rich source of additional information, including a detailed description of the life of an mRNA molecule, from transcription through intracellular trafficking, translation and degradation. Transcript imaging in live cells is considerably more challenging than in fixed cells, both because of the gentler chemical conditions that must be applied to preserve cell integrity and the more intricate image processing steps, which require tracking the trajectory of a transcript with time, a task that is often confounded by the high diffusion rates of molecules and the proximity to other molecules[42]. Two main technologies enable real-time measurements of transcript levels in single cells: the MS2 method and molecular beacons.

The coat protein of the bacteriophage MS2 binds a specific RNA hairpin loop. In the MS2 system, a fluorescent reporter protein gene such as *GFP* is fused to the MS2 coat protein gene, and several successive target RNA hairpin–encoding sequences are inserted into the 3′ untranslated region of a target gene of interest. Simultaneous binding of several MS2-GFP proteins to the target transcript yields a concentrated fluorescence signal. This method has been used to follow the dynamics of gene expression in bacteria[43], yeast[44] and mammalian cells[45,46], achieving single-transcript resolution. Moreover, a general method to tag and image any mRNA in yeast has recently become available[47]. A limitation of the MS2-GFP approach is the high background fluorescence associated with the MS2-GFP molecules that are not bound to the target RNA. Split-GFP approaches[48–51] alleviate this limitation, as each fragment is attached to a different RNA-binding protein that targets a distinct RNA motif. By constructing these motifs in tandem at the 3′ untranslated region of the target, an intact fluorescent GFP is assembled only when bound to the target transcript. The main limitation of the MS2 and split-GFP methods is the requirement to generate transgenes for the *GFP* construct. Yoshio Umezawa and colleagues recently demonstrated the ability to target endogenous mRNA by fusing the GFP fragments to the RNA-binding domain of the Pumilio protein, which can be designed to facilitate specific binding to an arbitrary RNA sequence[51]. A remaining drawback is that the binding of multiple proteins to the target mRNA could modify its intracellular dynamics, thus providing a nonrepresentative picture of the endogenous conditions.

Another technique for following individual transcripts in live cells uses molecular beacons[52,53]—nucleic acid probes that form hairpins coupled to a fluorophore and a quencher on opposite ends. Specific hybridization to the target sequence causes a conformational change that physically separates the quencher and fluorophore, resulting in light emission. Transcript detection in live cells is more challenging owing to their decreased stability caused by intracellular RNases and their non-homogenous intracellular distribution[49]. Probes with 2′-O-methylribonucleotide, which is not a target of RNase H, have been shown to alleviate this problem[54–56]. Molecular beacons require microinjection into cells, a procedure with high yield but one that may affect cell viability and can often result in a rapid drift of the microinjected probes into the nucleus. As an alternative, Santangelo *et al.*[56] used reversible permeabilization of the plasma membrane with pore-forming toxins, such as streptolysin-O, to deliver multiply labeled tetravalent RNA probes with minimal cytotoxicity. These probes, which are complementary to the target RNA and labeled with three fluorescent molecules each, tetramerize through their additional binding to streptavidin, thus yielding increased accumulation of fluorophores at the target transcript and allowing visualization of RNA dynamics[56].

Imaging technology for single-molecule FISH

Detection of single mRNA molecules imposes specific requirements on the imaging platform in terms of spatial resolution and sensitivity. For lens-based systems the maximum resolution can be estimated by the Rayleigh criteria, which says the resolution is equal to $(0.61)\lambda$ / NA, in which λ is the illumination wavelength, and NA is the numerical aperture of the lens. In practice, this limits the lateral resolution to 200–400 nm. Optimizing spatial resolution typically requires the use of a high-numerical-aperture oil-immersion lens and immersion oil that exactly matches the refractive index of the lens and the cover glass. Achromat or apochromat objectives minimize chromatic aberrations and are thus necessary when multiplexing different fluorophores[42]. Resolving diffraction-limited volumes also requires using a camera with sufficient spatial resolution that does not compromise the optimal optical resolution. CCD cameras have pixel sizes of 2–40 μm, which translates to a size of 20–400 nm at the image plane with a 100× objective depending on the optical setup. To achieve optimal resolution, it is crucial to determine the size of the image being projected onto the camera and verify that there are 2.5–3 pixels per unit resolution. If pixel binning is used to increase the signal at the camera, the number of required pixels must be multiplied accordingly. Exposure times should be optimized: long exposure times lead to higher precision in transcript localization but could accelerate sample bleaching[57].

Image signal is often masked by noise stemming from both instrument-related factors such as dark current, pixelation noise and CCD readout noise but mostly from out-of-focus light from cellular autofluorescence[57,58]. Maximizing signal-to-noise ratio requires increasing photon output of the fluorescent labels and limiting background noise. Wide-field epi-illuminated microscopes use a broad-band light source with a spectrum from 250 nm to 1,100 nm. Excitation filters can narrow this band to the fluorophore excitation wavelength, but autofluorescence always remains. Collecting light from a wide field also results in an increase in background. Confocal microscopes have the advantage of using a pinhole or slit to block detection of light from outside the focal region, thus potentially increasing relevant fluorescence and decreasing autofluorescence. Disadvantages of confocal microscopes over epi-fluorescence microscopes include their higher cost, the lower throughput owing to the necessity to scan the image (partly alleviated by using spinning disk confocal microscopes), more stringent safety requirements and the additional descanning optics necessary to guide the detected light to the detector, which can limit transmission efficiency. Multiplex detection of several fluorophores can be achieved by using filter wheels for epifluorescence microscopes and multimode lasers for confocal microscopes. Instrument-related noise, such as dark current can be substantially minimized by cooling the CCD chip to −80 °C (ref. 58).

Data analysis

Extracting gene expression from single-molecule FISH experiments requires automatic detection of spots representing single

© 2011 Nature America, Inc. All rights reserved.

mRNA molecules in digital three-dimensional image stacks. The image processing steps usually include image enhancement—either deconvolution or high-pass filtering, followed by thresholding and localization of connected components[21,22,28,31,34]. A point source in the illuminated image plane is expanded and distorted by the microscope and the detector optics, resulting in a blurred three-dimensional version, called the point spread function (PSF). Because a fluorescence emission pattern can be treated as a sum of point sources, a recorded image is a convolution of the real image and the PSF. When an accurate description of the PSF is available, one can use deconvolution algorithms to reconstruct the original image[59]. Such algorithms deblur images and reassign photons emanating from out-of-focus *z*-dimension planes back to their original plane. Successful deconvolution requires precise measurements of the PSF, which is typically achieved by imaging sub-diffraction fluorescent beads[21,22]. An alternative to deconvolution is to filter the image stack with a three-dimensional band-pass filter (such as a three-dimensional Laplacian of Gaussian filter) to enhance features of the relevant spot dimension[28,31]. The step after image enhancement usually involves applying a threshold to the enhanced image to yield a binary image in which connected components can be localized. Because a connected component spans more than a single pixel, its centroid achieves subpixel resolution for source localization. The value of the image thresholds can affect the number of spots detected, and a reasonable choice for this value is one at which the number of detected spots is least sensitive to threshold selection[28,31]. Alternatively, least-square Gaussian fitting algorithms can be used to localize the point source in a grayscale image[57,60].

Outlook

One appealing application of single-transcript imaging is the validation of regulatory interactions. A traditional approach entails engineering mice in which a regulatory gene of interest is deleted and searching for putative target genes with modified expression[61]. Single-molecule FISH can enable detection of such targets in wild-type tissue by hybridizing to a sample probes for the regulatory gene and for a putative target gene. A measured positive correlation in the transcript abundance could imply either a direct regulation or, alternatively, regulation by a common upstream component. Single-molecule transcript imaging can also provide valuable information on the behavior of network motifs, modular circuit components such as feedback and feed-forward loops, which are highly abundant in transcriptional networks and often comprise only a few genes[62]. Simultaneous quantitative *in situ* measurements of a handful of different transcripts can shed light on the behavior of these motifs in their tissue context. In a tumor, single-molecule transcript imaging can highlight the role of transcriptional heterogeneity in tumor progression and the relation between spatial context and phenotypic states of cells, represented by their expression signatures.

Single-molecule transcript imaging techniques can be combined with high-throughput expression analysis in two complementary ways. One would be to start out with large gene expression screens that would suggest putative genes of interest, the detailed *in situ* expression of which would then be described using single-molecule FISH. An alternative approach would be to start with an unbiased mapping of a tissue using single-molecule FISH probes to detect an interesting expression pattern in terms of spatial distribution in a tissue or an unusual co-expression pattern of a few genes in isolated cells. One could then enrich for such cells using FACS or laser capture and extend this core gene expression signature with high-throughput genome-wide expression measurements. This approach could provide a detailed description of rare cell populations in a tumor, such as putative cancer stem cells.

A technical limitation of single-molecule transcript imaging is the inability to spatially resolve single molecules when they are closer than the diffraction limit, typically 200 nm. This could be a considerable problem for highly expressed genes such as ribosomal components, especially in smaller organisms such as yeast[25] and when mRNA molecules are physically localized in transport particles[63]. Techniques that can be used to address this limitation are sub–diffraction-limit microscopy methods such as stimulated emission depletion microscopy[64], stochastic optical reconstruction microscopy[65] and photoactivated localization microscopy[66], which enable resolving fluorescent molecules with nanometer resolution[67]. Subdiffraction microscopy outperforms other technologies in terms of spatial resolution, enabling probing molecular structures in fixed and even live cells[68,69], but scaling the technology to comprehensively measure gene expression *in situ* in many cells and in tissues is still a challenge because of long recording times, high-intensity illumination, the prolonged use of which could potentially be harmful to the sample, as well as expensive instrumentation.

Another exciting recent development is the use of probes coupled to quantum dots[70–73]. The brightness of quantum dots makes them especially attractive for studying tissues, in which cellular autofluorescence often masks the signal. Oligonucleotide probes labeled with quantum dots have been used to detect transcripts in cells and tissues[70–72], in paraffin-embedded tissue[74] and even in live-cell imaging[75]. Some limitations of using quantum dots, such as reduced permeability and steric hindrance difficulties when binding the targets, mainly caused by their relatively large size compared to conventional probes, as well as their tendency to turn on and off ('blinking')[76] still limit the use of quantum dot–labeled probes for transcript imaging but their attractive photophysical properties suggest a huge potential for single-molecule detection.

Finally, a challenge for the future is to combine transcript-imaging approaches with quantitative measurements of other cellular constituents, namely proteins and DNA. While protocols tailored to simultaneously perform RNA FISH and immunofluorescence have been successful in some cases[56,69], their generic use is still limited by the variability inherent in immunofluorescence. Combination of RNA FISH with immunofluorescence or with GFP signal measurements will facilitate decoupling the relative contributions of transcriptional and translational regulation in cells and tissues, whereas combination with DNA FISH can address the expression variability of different clones in a tumor tissue. Such analysis can provide important insights into the combined regulation of protein expression in complex tissues.

ACKNOWLEDGMENTS

We thank S. Semrau, J.P. Junker, S. Mukherji and A. Lavi-Itzkovitz for valuable comments. This work was supported by the US National Institutes of Health National Cancer Institute Physical Sciences Oncology Center at the Massachusetts Institute of Technology (U54CA143874) and a US National Institutes of Health Pioneer award (1DP1OD003936) to A.v.O.;

© 2011 Nature America, Inc. All rights reserved.
npg

S.I. acknowledges support from the European Molecular Biology Organization, the Human Frontiers Science Program and the Machiah Foundation.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

© 2011 Nature America, Inc. All rights reserved.

1. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
2. Heller, R.A. *et al.* Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. USA* **94**, 2150–2155 (1997).
3. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869–10874 (2001).
4. van't Veer, L.J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
5. Levsky, J.M. & Singer, R.H. Gene expression and the myth of the average cell. *Trends Cell Biol.* **13**, 4–6 (2003).
6. Raj, A. & van Oudenaarden, A. Nature, nurture or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
7. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).
8. Park, S.Y., Gonen, M., Kim, H.J., Michor, F. & Polyak, K. Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *J. Clin. Invest.* **120**, 636–644 (2010).
9. Navin, N. *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20**, 68–80 (2010).
10. Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. Stem cells, cancer and cancer stem cells. *Nature* **414**, 105–111 (2001).
11. Emmert-Buck, M.R. *et al.* Laser capture microdissection. *Science* **274**, 998–1001 (1996).
12. Freeman, W.M., Walker, S.J. & Vrana, K.E. Quantitative RT-PCR: pitfalls and potential. *Biotechniques* **26**, 112–125 (1999).
13. Warren, L., Bryder, D., Weissman, I.L. & Quake, S.R. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc. Natl. Acad. Sci. USA* **103**, 17807–17812 (2006).
14. van der Flier, L.G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–912 (2009).
15. Levsky, J.M. & Singer, R.H. Fluorescence *in situ* hybridization: past, present and future. *J. Cell Sci.* **116**, 2833–2838 (2003).
16. van der Ploeg, M. Cytochemical nucleic acid research during the twentieth century. *Eur. J. Histochem.* **44**, 7–42 (2000).
17. Gregorieff, A. & Clevers, H. *In situ* hybridization to identify gut stem cells. *Curr. Protoc. Stem. Cell. Biol.* chapter 11, unit 2F.1 (2010).
18. Pare, A. *et al.* Visualization of individual *Scr* mRNAs during *Drosophila* embryogenesis yields evidence for transcriptional bursting. *Curr. Biol.* **19**, 2037–2042 (2009).
19. Randolph, J.B. & Waggoner, A.S. Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. *Nucleic Acids Res.* **25**, 2923–2929 (1997).
20. Larson, D.R., Singer, R.H. & Zenklusen, D. A single molecule view of gene expression. *Trends Cell Biol.* **19**, 630–637 (2009).
21. Femino, A.M., Fay, F.S., Fogarty, K. & Singer, R.H. Visualization of single RNA transcripts *in situ*. *Science* **280**, 585–590 (1998).
This pioneering work demonstrated single-molecule transcript imaging in fixed cells using multiply-labeled fluorescent probes.
22. Femino, A.M., Fogarty, K., Lifshitz, L.M., Carrington, W. & Singer, R.H. Visualization of single molecules of mRNA *in situ*. *Methods Enzymol.* **361**, 245–304 (2003).
23. Maamar, H., Raj, A. & Dubnau, D. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* **317**, 526–529 (2007).
24. Zenklusen, D., Larson, D.R. & Singer, R.H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15**, 1263–1271 (2008).
25. Tan, R.Z. & van Oudenaarden, A. Transcript counting in single cells reveals dynamics of rDNA transcription. *Mol. Syst. Biol.* **6**, 358 (2010).
26. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
27. Capodieci, P. *et al.* Gene expression profiling in single cells within tissue. *Nat. Methods* **2**, 663–665 (2005).
28. 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).
Single-molecule transcript imaging in cells and in fixed tissues was achieved using multiple singly labeled probes.
29. To, T.L. & Maheshri, N. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* **327**, 1142–1145 (2010).
30. Khalil, A.M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 11667–11672 (2009).
31. Raj, A., Rifkin, S.A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).
32. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
33. Lu, J. & Tsourkas, A. Imaging individual microRNAs in single mammalian cells *in situ*. *Nucleic Acids Res.* **37**, e100 (2009).
34. Larsson, C., Grundberg, I., Soderberg, O. & Nilsson, M. *In situ* detection and genotyping of individual mRNA molecules. *Nat. Methods* **7**, 395–397 (2010).
This work introduced a sensitive single-molecule transcript-imaging technique that uses padlock probes and *in situ* target-primed rolling-circle amplification in cells and tissues to specifically detect targets with a single nucleotide difference.
35. Wittung, P., Nielsen, P.E., Buchardt, O., Egholm, M. & Norden, B. DNA-like double helix formed by peptide nucleic acid. *Nature* **368**, 561–563 (1994).
36. Lansdorp, P.M. *et al.* Heterogeneity in telomere length of human chromosomes. *Hum. Mol. Genet.* **5**, 685–691 (1996).
37. Kloosterman, W.P., Wienholds, E., de Bruijn, E., Kauppinen, S. & Plasterk, R.H. *In situ* detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nat. Methods* **3**, 27–29 (2006).
38. Player, A.N., Shen, L.P., Kenny, D., Antao, V.P. & Kolberg, J.A. Single-copy gene detection using branched DNA (bDNA) *in situ* hybridization. *J. Histochem. Cytochem.* **49**, 603–612 (2001).
39. Levsky, J.M., Shenoy, S.M., Pezo, R.C. & Singer, R.H. Single-cell gene-expression profiling. *Science* **297**, 836–840 (2002).
This work substantially increased the number of targets that can be simultaneously imaged by using combination of probes labeled with spectrally distinct fluorophores.
40. Nederlof, P.M., van der Flier, S., Vrolijk, J., Tanke, H.J. & Raap, A.K. Fluorescence ratio measurements of double-labeled probes for multiple *in situ* hybridization by digital imaging microscopy. *Cytometry* **13**, 839–845 (1992).
41. Nederlof, P.M. *et al.* Multiple fluorescence *in situ* hybridization. *Cytometry* **11**, 126–131 (1990).
42. Park, H.Y., Buxbaum, A.R. & Singer, R.H. Single mRNA tracking in live cells. *Methods Enzymol.* **472**, 387–406 (2010).
43. Golding, I., Paulsson, J., Zawilski, S.M. & Cox, E.C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
44. Bertrand, E. *et al.* Localization of *ASH1* mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
This work introduced the MS2 method for single mRNA detection in live cells.
45. Fusco, D. *et al.* Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.* **13**, 161–167 (2003).
46. Shav-Tal, Y. *et al.* Dynamics of single mRNPs in nuclei of living cells. *Science* **304**, 1797–1800 (2004).
47. Haim, L., Zipor, G., Aronov, S. & Gerst, J.E. A genomic integration method to visualize localization of endogenous mRNAs in living yeast. *Nat. Methods* **4**, 409–412 (2007).
48. Rackham, O. & Brown, C.M. Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J.* **23**, 3346–3355 (2004).
49. Tyagi, S. Imaging intracellular RNA distribution and dynamics in living cells. *Nat. Methods* **6**, 331–338 (2009).
50. Valencia-Burton, M., McCullough, R.M., Cantor, C.R. & Broude, N.E. RNA visualization in live bacterial cells using fluorescent protein complementation. *Nat. Methods* **4**, 421–427 (2007).
51. Ozawa, T., Natori, Y., Sato, M. & Umezawa, Y. Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat. Methods* **4**, 413–419 (2007).

52. Tyagi, S. & Kramer, F.R. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* **14**, 303–308 (1996). **Molecular beacons were introduced for the detection of single mRNAs in live cells.**
53. Vargas, D.Y., Raj, A., Marras, S.A., Kramer, F.R. & Tyagi, S. Mechanism of mRNA transport in the nucleus. *Proc. Natl. Acad. Sci. USA* **102**, 17008–17013 (2005).
54. Bratu, D.P., Cha, B.J., Mhlanga, M.M., Kramer, F.R. & Tyagi, S. Visualizing the distribution and transport of mRNAs in living cells. *Proc. Natl. Acad. Sci. USA* **100**, 13308–13313 (2003).
55. Lifland, A.W., Zurla, C. & Santangelo, P.J. Single molecule densitive multivalent polyethylene glycol probes for RNA imaging. *Bioconjug. Chem.* **21**, 483–488 (2010).
56. Santangelo, P.J. *et al.* Single molecule-sensitive probes for imaging RNA in live cells. *Nat. Methods* **6**, 347–349 (2009).
57. Thompson, R.E., Larson, D.R. & Webb, W.W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775–2783 (2002).
58. Grunwald, D., Singer, R.H. & Czaplinski, K. Cell biology of mRNA decay. *Methods Enzymol.* **448**, 553–577 (2008).
59. McNally, J.G., Karpova, T., Cooper, J. & Conchello, J.A. Three-dimensional imaging by deconvolution microscopy. *Methods* **19**, 373–385 (1999).
60. Smith, C.S., Joseph, N., Rieger, B. & Lidke, K.A. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. Methods* **7**, 373–375 (2010).
61. van der Flier, L.G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu. Rev. Physiol.* **71**, 241–260 (2009).
62. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
63. Bassell, G. & Singer, R.H. mRNA and cytoskeletal filaments. *Curr. Opin. Cell Biol.* **9**, 109–115 (1997).
64. Klar, T.A., Jakobs, S., Dyba, M., Egner, A. & Hell, S.W. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl. Acad. Sci. USA* **97**, 8206–8210 (2000).
65. Rust, M.J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).
66. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
67. Hell, S.W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
68. Shroff, H., Galbraith, C.G., Galbraith, J.A. & Betzig, E. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nat. Methods* **5**, 417–423 (2008).
69. Bassell, G.J. *et al.* Sorting of beta-actin mRNA and protein to neurites and growth cones in culture. *J. Neurosci.* **18**, 251–265 (1998).
70. Choi, Y. *et al. In situ* visualization of gene expression using polymer-coated quantum-dot-DNA conjugates. *Small* **5**, 2085–2091 (2009).
71. Chan, P., Yuen, T., Ruf, F., Gonzalez-Maeso, J. & Sealfon, S.C. Method for multiplex cellular detection of mRNAs using quantum dot fluorescent *in situ* hybridization. *Nucleic Acids Res.* **33**, e161 (2005).
72. Gonzalez-Maeso, J. *et al.* Identification of a serotonin/glutamate receptor complex implicated in psychosis. *Nature* **452**, 93–97 (2008).
73. Smith, A.M., Duan, H., Mohs, A.M. & Nie, S. Bioconjugated quantum dots for in vivo molecular and cellular imaging. *Adv. Drug Deliv. Rev.* **60**, 1226–1240 (2008).
74. Tholouli, E. *et al.* Imaging of multiple mRNA targets using quantum dot based in situ hybridization and spectral deconvolution in clinical biopsies. *Biochem. Biophys. Res. Commun.* **348**, 628–636 (2006).
75. Ishihama, Y. & Funatsu, T. Single molecule tracking of quantum dot-labeled mRNAs in a cell nucleus. *Biochem. Biophys. Res. Commun.* **381**, 33–38 (2009).
76. Weil, T.T., Parton, R.M. & Davis, I. Making the message clear: visualizing mRNA localization. *Trends Cell Biol.* **20**, 380–390 (2010).
77. Tautz, D. & Pfeifle, C. A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* **98**, 81–85 (1989).
78. Bloom, K.S. *et al.* Using green fluorescent protein fusion proteins to quantitate microtubule and spindle dynamics in budding yeast. *Methods Cell Biol.* **61**, 369–383 (1999).

© 2011 Nature America, Inc. All rights reserved.