Imaging individual mRNA molecules using multiple singly labeled probes

Arjun Raj[1], Patrick van den Bogaard[2], Scott A Rifkin[1], Alexander van Oudenaarden[1] & Sanjay Tyagi[2]

We describe a method for imaging individual mRNA molecules in fixed cells by probing each mRNA species with 48 or more short, singly labeled oligonucleotide probes. This makes each mRNA molecule visible as a computationally identifiable fluorescent spot by fluorescence microscopy. We demonstrate simultaneous detection of three mRNA species in single cells and mRNA detection in yeast, nematodes, fruit fly wing discs, and mammalian cell lines and neurons.

As it is becoming increasingly apparent that gene expression in individual cells deviates substantially from the average behavior of cell populations[1], new methods that provide accurate integer counts of mRNA copy numbers in individual cells are needed. Ideally, such methods should also reveal the intracellular locations of the mRNAs, as mRNA localization is often used by cells to spatially restrict the activity of proteins[2]. One candidate for such a method is *in situ* hybridization followed by microscopic analysis[3,4]. A conventional practice is to link probes to enzymes that catalyze chromogenic or fluorogenic reactions[5]. However, because the products of these reactions are small molecules or precipitates that diffuse away from the probe, the location of the target molecule is not precisely determined. Conversely, probes labeled directly with a few fluorophores maintain spatial resolution, but the sensitivity that can be achieved is relatively poor.

To circumvent these problems, a variant of fluorescence *in situ* hybridization (FISH) procedure has been developed that is sensitive enough to detect single mRNA molecules[6]. In this procedure, 5 oligonucleotide probes, each about 50 nucleotides long and labeled with 5 fluorophore moieties, are hybridized to each mRNA target, which then becomes visible as a diffraction-limited fluorescent spot. Although these probes have been used successfully[7], the system has not been widely adopted. One reason for this is difficulty in synthesizing and purifying heavily labeled oligonucleotides: the amine groups used for coupling fluorophores to the probe are prone to loss, and it is hard to purify fully coupled probes from partially coupled ones[8]. Also, when some fluorophores are present in multiple copies on the same oligonucleotide, they interact with each other, altering the hybridization characteristics of the oligonucleotides and resulting in severe self-quenching[9].

Another issue with the use of small numbers of heavily labeled probes is that the signals are more prone to variability. For instance, when using 5 fluorescent probes targeted to a single mRNA, the researchers had estimated that the majority of the fluorescent spots observed have intensities corresponding to the presence of only 1 or 2 probes[6]. This makes it difficult to unambiguously identify all the fluorescent spots as mRNA molecules as it is impossible to determine whether the detection of an individual probe arises from legitimate binding to the target mRNA or from nonspecific binding.

To address these issues, we reasoned that by taking advantage of the high throughput of 96-position DNA synthesizers, one could synthesize a large number of probes and reliably label them with a single fluorophore moiety at their 3′ termini to detect individual mRNA molecules. We constructed a doxycycline-controlled gene

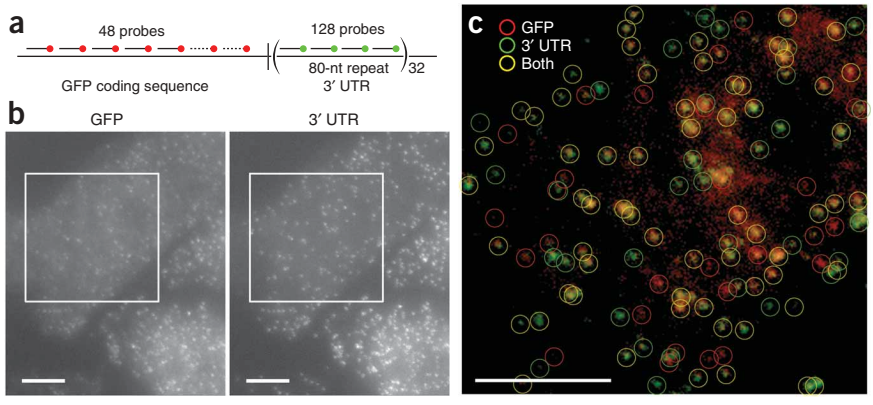

Figure 1 | Simultaneous detection of a unique sequence and a repeated sequence in individual mRNA molecules. (**a**) Schematic of the construct used. The 48 probes used to detect the GFP coding sequence were labeled with Alexa 594, and the four different probes used to detect the tandem repeat in the 3′ UTR were labeled with TMR. (**b**) Maximum intensity merges of a pair of *z*-image stacks of fluorescent images of CHO cells taken in the Alexa 594 channel (left) and the TMR channel (right). (**c**) False-color merge of the boxed regions in **b**, with circles representing computationally identified mRNA particles. Scale bars, 5 μm.

[1]Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [2]Public Health Research Institute Center, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, 225 Warren Street, Newark, New Jersey 07103, USA. Correspondence should be addressed to A.R. (arjunraj@mit.edu) or S.T. (tyagisa@umdnj.edu).

RECEIVED 25 JUNE; ACCEPTED 25 AUGUST; PUBLISHED ONLINE 21 SEPTEMBER 2008; DOI:10.1038/NMETH.1253

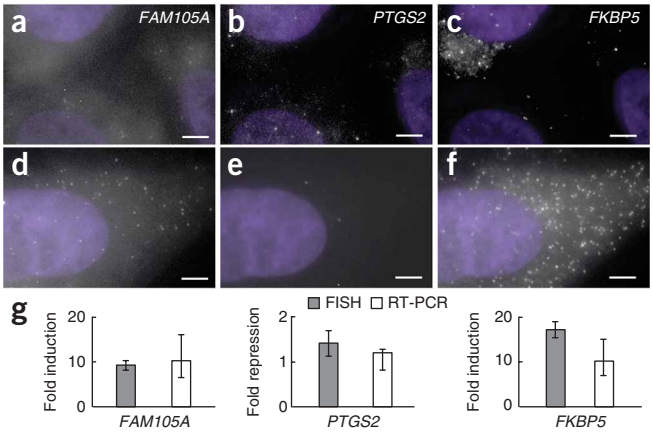

Figure 2 | Simultaneous imaging of three different mRNAs in mammalian cells. (**a–f**) Images showing fluorescent spots corresponding to *FAM105A*, *PTGS2* and *FKBP5* mRNAs in the same set of A549 cells not treated with dexamethasone (**a–c**) and in cells treated for 8 h with 24 nM dexamethasone (**d–f**). (**g**) Fold change in expression for all three genes as measured by FISH and real-time RT-PCR; error bars represent standard errors of measurements (see **Supplementary Methods**). All images are maximum merges of a *z*-stack of fluorescence images spanning the extent of the cells with nuclear 4,6-diamidino-2-phenylindole (DAPI) counterstaining in purple. Scale bars, 5 μm.

that produced an mRNA encoding GFP and contained 32 tandemly repeated 80-nucleotide-long sequences in its 3′ untranslated region (UTR); we then stably integrated this engineered gene into the genome of a Chinese hamster ovary cell line (**Fig. 1a**). Previously, we have shown that fluorescent probes targeted to tandemly repeated copies of probe-binding sequence results in FISH signals corresponding to individual molecules using a variety of methods, including a demonstration that the number of fluorescent spots per cell was about the same as the number of mRNA per cell, as measured by quantitative real-time reverse-transcriptase (RT)-PCR[10,11]. Here we targeted the coding region of the *GFP* mRNA with 48 different oligonucleotides labeled with Alexa 594 fluorophores and targeted each repeat sequence in the 3′ UTR with 4 oligonucleotides labeled with tetramethylrhodamine (TMR).

After hybridization, we imaged the cells with a pair of filter sets that could clearly distinguish between the two fluorophores. We found many 'particles' with a diameter of about 0.25 μm that appeared in both the TMR and Alexa 594 channels (**Fig. 1b**). The particles were identified computationally using an image processing program (**Supplementary Fig. 1**, **Supplementary Methods** and **Supplementary Software** online) that categorizes particles as being labeled with either the GFP-coding-sequence probes (TMR), the UTR-specific probes (Alexa-594) or both (**Fig. 1c**). Upon identifying and localizing particles in four fields of view similar to the ones shown in **Figure 1c**, we counted a total of 599 particles corresponding to GFP coding sequence–specific probes and 565 particles corresponding to the UTR-specific probes. Of these particles, 85% of the 'UTR particles' localized with the 'GFP particles', whereas 81% of the GFP particles colocalized with the UTR particles. The high degree of colocalization between particles detected by the previously established tandem-repeat detection method[10] and the particles detected via simultaneous probing with 48 different singly labeled oligonucleotides demonstrates the validity of using multiple single-labeled probes for the detection of endogenous transcripts. The fraction of particles that did not colocalize likely corresponded to mRNA molecules that lost either their coding sequence or their 3′ UTR via the natural processes of mRNA degradation. An analysis of fluorescence intensity of the colocalized spots showed that the spot intensities displayed a unimodal distribution (**Supplementary Fig. 2** online), arguing that the particles detected were individual molecules[10].

We also explored how the signal intensity varied with the number of probes by performing FISH using either the first 12, 24, or 36 probes or all 48 probes in our set. For this particular target mRNA, we found that particles could be detected with fewer probes, albeit with decreased intensity (**Supplementary Fig. 3a** online). However, our automatic spot-detection algorithm performed particularly well with 48 probes, detecting the same number of spots over a broad range of thresholds (**Supplementary Fig. 3b**). The number of probes required for robust signal is likely to depend on the target sequence, though, as accessibility to probes depends on the secondary structure in the RNA. Our method was at least as sensitive as the FISH-based method[6] described above (**Supplementary Fig. 4** online).

A potential use of our method is to simultaneously detect single molecules of multiple mRNAs in individual cells. To detect three different mRNAs at the same time, we designed probes specific for mRNAs encoding *FKBP5*, *PTGS2* and *FAM105A* in the human carcinoma cell line A549. We coupled these probes to the spectrally distinct fluorophores Cy5, Alexa 594 and TMR, respectively. Upon performing FISH with all three probes simultaneously, individual spots were visible in the three different fluorescence channels (**Fig. 2a–f**). The spots corresponding to different mRNAs did not overlap with each other. An intensity analysis showed that fluorescent spots did not bleed through into other channels (**Supplementary Fig. 5** online) and the use of an oxygen-scavenging mounting buffer ensured the stability of all fluorophores during the acquisition of image stacks (**Supplementary Fig. 6** online).

To demonstrate that our method of mRNA detection was specific and quantitative, we added to the growth medium a cell-permeant glucocorticoid, dexamethasone, which upregulates expression of *FKBP5* and *FAM105A*, and mildly downregulates expression of *PTGS2* in this cell line[12]. The mean number of *FKBP5* and *FAM105A* mRNAs measured by combining FISH with our spot-detection algorithm increased whereas the mean number of *PTGS2* mRNAs decreased (**Fig. 2a–f**). The values we obtained corresponded well to RT-PCR measurements of the fold induction and repression of these genes performed on the same samples, demonstrating that the fluorescent spots are the appropriate mRNAs and that we detected a majority of the mRNA molecules (**Fig. 2g**). This also demonstrated the effectiveness of our spot detection algorithm for accurate gene-expression quantification.

Our method also captured spatial information about the location of the mRNAs detected, a particularly important feature for studying development, in which mRNAs often display spatial patterning. We tested our method for efficacy in two commonly studied developmental systems: the nematode, *Caenorhabditis elegans*, and the fruit fly, *Drosophila melanogaster*. In the nematode, we constructed probes to detect mRNA molecules transcribed from the gene *elt-2*, a transcription factor that is expressed only in the nematode gut and only after the embryo has developed to the

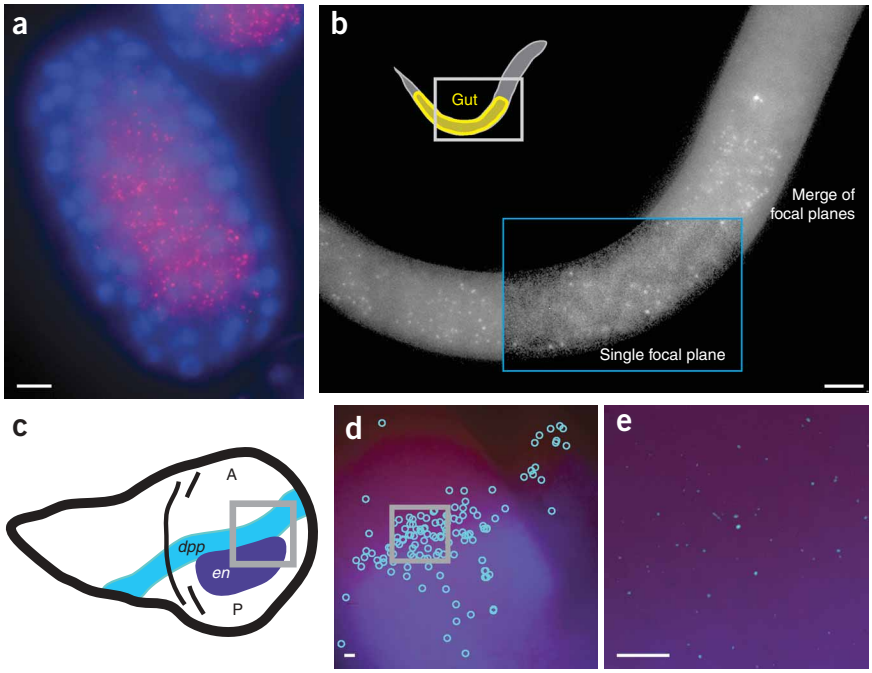

Figure 3 | Imaging localized mRNAs in *C. elegans* and *D. melanogaster*. (**a**) *elt-2* mRNA molecules (magenta) in an early stage *C. elegans* embryo (~100 cell stage); the nuclei were counterstained with DAPI (blue). (**b**) *elt-2* mRNA molecules in a *C. elegans* L1 larva. A single focal plane is shown in the boxed region, in which the intestinal track is visible. (**c**) A schematic of *dpp* and engrailed (*en*) expression in the imaginal wing discs of third instar *D. melanogaster* larvae. (**d**) Image showing the locations of the computationally identified *dpp* mRNA molecules (light blue circles) and Engrailed expression detected by immunofluorescence (dark blue). (**e**) Magnification of the boxed region in **d** showing enhanced *dpp* mRNA molecule signals (light blue) and Engrailed protein expression detected by immunofluorescence (dark blue). All images except the boxed portion in **b** are maximum merges of a *z*-stack of fluorescence images. Scale bars, 5 μm.

45-cell stage[13]. After hybridization of the probe set to both embryos and larvae, we found that *elt-2* mRNA molecules were present only in the gut region (**Fig. 3a**) of both the embryos and the larvae (**Fig. 3b**). Consistent with the known timing of the onset of expression[13], we only detected *elt-2* mRNAs in the gut of embryos older than the 45-cell stage.

In the fruit fly, one of the most well-studied examples of the localization of gene expression occurs in wing imaginal disc development[14]. The wing discs of fruit fly larvae display a remarkable set of gene expression patterns, one of which is the formation of a stripe of expression of the gene *dpp* in response to gradients of the morphogenic proteins Hedgehog and Engrailed[14] (**Fig. 3c**). To check whether this narrow stripe of *dpp* mRNA synthesis can be imaged, we constructed a set of singly labeled probes against *dpp* mRNA and performed FISH on imaginal wing discs isolated from third instar larvae while simultaneously performing immunofluorescence against Engrailed protein. We detected *dpp* mRNA in a stripe along the boundary of Engrailed protein expression (**Fig. 3d,e**), demonstrating both that the method can be used in wing imaginal discs and that the method can be easily combined with immunofluorescence detection.

Additional tests of our method showed that it was also applicable to *Saccharomyces cerevisae* and cultured rat hippocampal neurons, showing expected specificity in salt-induced expression of the *STL1* gene and dendritic localization of β-actin and *MAP2* mRNAs (**Supplementary Fig. 7** online).

Here we described a FISH method that allows for multiplex gene-expression profiling of transcripts across many model organisms. By using large numbers of singly labeled probes, our method generates uniform signals that can be computationally identified to yield accurate mRNA counts. In contrast, methods using heavily labeled probes (such as dendrimers) can suffer from false positives and negatives owing to individual probe misbinding or nonbinding events, respectively. Another advantage is the simplicity of probe generation and purification; by pooling, coupling and purifying the probes *en masse*, much of the complexity of probe preparation can be avoided. We created a web-based program for designing probe sets with optimally uniform G+C content (http://www.singlemoleculefish.com/). The simplicity of our method will likely facilitate genomic-scale studies of mRNA number and localization with applications in systems biology, cell biology, neurobiology and developmental biology.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank S.A.E. Marras, F.R. Kramer, D. Vargas, S. Sinha, E. Andersen, G. Neuert and Q. Yang. This work was supported by the US National Institute of Mental Health grant MH079197; National Institute of General Medicine grants GM068957, GM077183 and GM070357; and National Science Foundation grant PHY0548484. A.R. is supported by a National Science Foundation fellowship DMS-0603392 and a Burroughs Wellcome Fund Career Award at the Scientific Interface.

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

1. Kaufmann, B.B. & van Oudenaarden, A. *Curr. Opin. Genet. Dev.* **17**, 107–112 (2007).
2. St. Johnston, D. *Nat. Rev. Mol. Cell Biol.* **6**, 363–375 (2005).
3. Gall, J.G. *Proc. Natl. Acad. Sci. USA* **60**, 553–560 (1968).
4. Levsky, J.M. & Singer, R.H. *J. Cell Sci.* **116**, 2833–2838 (2003).
5. Raap, A.K. *et al. Hum. Mol. Genet.* **4**, 529–534 (1995).
6. Femino, A.M., Fay, F.S., Fogarty, K. & Singer, R.H. *Science* **280**, 585–590 (1998).
7. Maamar, H., Raj, A. & Dubnau, D. *Science* **317**, 526–529 (2007).
8. Femino, A.M., Fogarty, K., Lifshitz, L.M., Carrington, W. & Singer, R.H. *Methods Enzymol.* **361**, 245–304 (2003).
9. Randolph, J.B. & Waggoner, A.S. *Nucleic Acids Res.* **25**, 2923–2929 (1997).
10. Vargas, D.Y., Raj, A., Marras, S.A., Kramer, F.R. & Tyagi, S. *Proc. Natl. Acad. Sci. USA* **102**, 17008–17013 (2005).
11. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. & Tyagi, S. *PLoS Biol.* **4**, e309 (2006).
12. Wang, J.C. *et al. Proc. Natl. Acad. Sci. USA* **101**, 15603–15608 (2004).
13. Fukushige, T., Hawkins, M.G. & McGhee, J.D. *Dev. Biol.* **198**, 286–302 (1998).
14. Sanicola, M., Sekelsky, J., Elson, S. & Gelbart, W.M. *Genetics* **139**, 745–756 (1995).