Dense transcript profiling in single cells by image correlation decoding

Ahmet F Coskun[1,2] & Long Cai[1,2]

Sequential barcoded fluorescent *in situ* hybridization (seqFISH) allows large numbers of molecular species to be accurately detected in single cells, but multiplexing is limited by the density of barcoded objects. We present correlation FISH (corrFISH), a method to resolve dense temporal barcodes in sequential hybridization experiments. Using corrFISH, we quantified highly expressed ribosomal protein genes in single cultured cells and mouse thymus sections, revealing cell-type-specific gene expression.

Profiling molecules such as RNAs or proteins in cells is the key to exploring cell identities and can reveal patterns in gene regulatory networks. SeqFISH has enabled the detection of a large number of mRNA species in single cells[1], and a recent implementation using binary sequential barcodes demonstrated the multiplexing of 100–1,000 transcripts in single cells[2]. Exciting *in situ* sequencing approaches can also decode RNA species with nucleotide-level resolution[3,4]. However, barcoding a large number of transcripts can result in a high density of spots within the cell, which makes it difficult to resolve individual spots and identify barcodes across the hybridizations. We previously demonstrated that super-resolution microscopy improves the detection of abundant RNA molecules[5], and others have used advanced algorithms including compressed sensing to assist super-resolution microscopy[6,7]. However, a robust method to decode high-copy-number RNAs or other densely labeled molecules with conventional fluorescence microscopy is needed.

Here, we use image correlation to decode seqFISH temporal barcodes. In a simple example, we encode three RNA species such that each species only appears in a unique combination of two out of three total hybridization (Hyb) rounds (**Fig. 1a**). If a given RNA has a Hyb2 and Hyb3 code, only the FISH spots corresponding to that RNA will be detected in the same positions between Hyb2 and Hyb3; crosscorrelating Hyb2 and Hyb3 images will generate a positive correlation for that RNA, with an amplitude that is proportional to the number of molecules, while the other RNA species will not contribute to the crosscorrelation. Thus, the copy number of an RNA can be extracted from the crosscorrelation of the images corresponding to its barcode assignment even when the images are dense. We have made custom MATLAB source codes for corrFISH quantification (**Supplementary Software**) and test images available (https://github.com/singlecelllab/correlationFISH).

To test this approach, we simulated images corresponding to two rounds of hybridization with only one RNA in common between the two images (mimicking 30- × 30-μm cell size) and a ten-fold excess of uncorrelated RNAs (**Fig. 1b** and **Supplementary Fig. 1**). Crosscorrelation between Hyb1 and Hyb2 images[8–10] shows a distinct peak at the center (**Fig. 1b**) and provides the RNA copy number even when the spots overlap and no obvious colocalization occurs between the images. We previously observed that mRNAs follow a spatial Poisson distribution[5] in cells. As the variance of the Poisson distribution is equal to its mean, the peak value at the crosscorrelation is a good estimator for transcript copy number. In addition, given that the higher-order cumulants of the Poisson distribution are all equal to the mean, the higher-order crosscorrelation peaks are also good estimators for the corresponding barcode copy numbers. For simplicity, we implement only pairwise correlation for the proof-of-principle demonstrations.

The corrFISH counts scale linearly with increasing density, while the localization-based method of counting FISH spots yields significant errors in estimating RNA copy number at a density of more than 1 molecule/μm^2 (**Fig. 1c** and **Supplementary Fig. 2**). In simulated data, 100 A molecules in 30-μm cells were accurately measured in the presence of 10,000 B or C molecules with a coefficient of variation (CV) of 1.76 ± 1.54 (s.e.m., $n = 20$) (**Fig. 1d** and **Supplementary Fig. 3**). Thus, the copy number of A molecules, which comprise only 1% of total labeled RNA (at a density of more than 10 molecules/μm^2), can be estimated within two-fold of the mean value.

As an experimental demonstration, we used corrFISH to quantify highly abundant ribosomal protein transcripts in single cells. Mice embryos have a mosaic distribution of these transcripts in different tissue types[11]; thus, it would be interesting to profile their abundance at the single-cell level. Initially, we compared corrFISH with single-molecule FISH (smFISH) quantification[12] in the same cell. As a proof of concept, we quantified *Rps2* transcript in mammalian cell cultures in the presence of other dense transcripts using sequential hybridizations. In the first round, *Rps2* was hybridized with four other genes, and in the second round, *Rps2* was hybridized with a different set of four genes (**Fig. 1e** and **Supplementary Fig. 4**). In the third round, we only hybridized *Rps2*. This allowed us to perform two controls. First, we compared the detection of *Rps2* using corrFISH barcoding of Hyb1*Hyb2 to that from Hyb2*Hyb3, finding a good linear regression ($R = 0.83$; **Fig. 1f**). Next, we compared corrFISH results from Hyb1*Hyb2

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California, USA. [2]Program in Biochemistry and Molecular Biophysics, California Institute of Technology, Pasadena, California, USA. Correspondence should be addressed to L.C. (lcai@caltech.edu).

RECEIVED 22 FEBRUARY; ACCEPTED 11 MAY; PUBLISHED ONLINE 6 JUNE 2016; DOI:10.1038/NMETH.3895

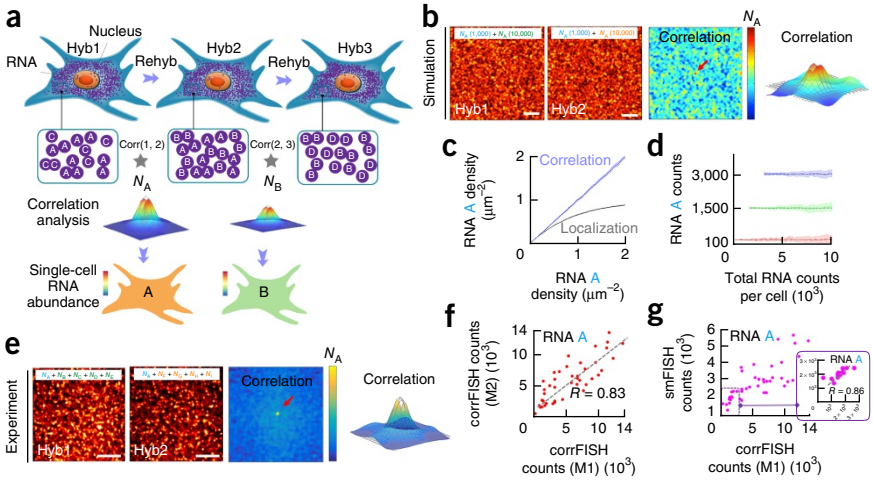

Figure 1 | Correlation FISH. (**a**) Schematic of sequential hybridizations of dense, barcoded mRNA molecules. RNA A is labeled in Hyb1 and Hyb2 and produces positive pairwise correlations across these two hybridization rounds but does not contribute to correlation between Hyb2 and Hyb3, in which only RNA B is labeled. Rehyb, rehybridization. Correlation peak values across Hybs (1,2) and (2,3) are estimators of the copy number of RNA A (N_A) and B (N_B), respectively, in that cell. Corr, correlation. (**b**) Simulated images of a cell with 1,000 A molecules in the presence of 10,000 B and C molecules in Hyb1 and Hyb2, respectively. The image correlation peak value (red arrow and 3D plot) corresponds to the copy number of A. Scale bars, 5 μm. (**c**) Correlation accurately estimates simulated RNA levels even at high density of RNAs, unlike localization-based dot counting. (**d**) Accuracy of detecting 100, 1,500, and 3,000 copies of A in the presence of increasing concentrations of B and C in simulated data (n = 20 simulation runs). Dashed line, expected number of A transcripts. (**e**) Experimental images of a cellular region with labeled transcripts of gene A (N_A) in the presence of four uncorrelated, labeled transcripts (N_{B-E}) and (N_{F-I}) in Hyb1 and Hyb2, respectively. The correlation functions yields transcript A abundance. Scale bars, 5 μm. (**f**) Gene A transcript detection in five-gene experiments using two different corrFISH barcodes (M1: Hyb1*Hyb2 and M2: Hyb2*Rps2 from Hyb3) with a linear fit of R = 0.83 and P < 0.002 (Student's t-test), n = 43 cells. (**g**) smFISH and corrFISH counts of transcript A are correlated. Inset, correlation R = 0.86 and P < 0.0006 (Student's t-test) in the linear regime, where smFISH quantification is not saturated.

to smFISH counting in Hyb3, yielding a correlation of R = 0.86 in the lower copy range (**Fig. 1g**). At high copy numbers (>3,000), smFISH quantification underestimates the transcript copy number because of FISH dot overlaps. These experiments show that corrFISH accurately quantifies high-copy-number genes in the presence of other high-copy-number transcripts. In fact, scaling this approach to all 79 ribosomal protein mRNAs requires similar densities of six genes per hybridization image (**Supplementary Fig. 5**).

To further test the performance of corrFISH, we targeted transcripts for ten

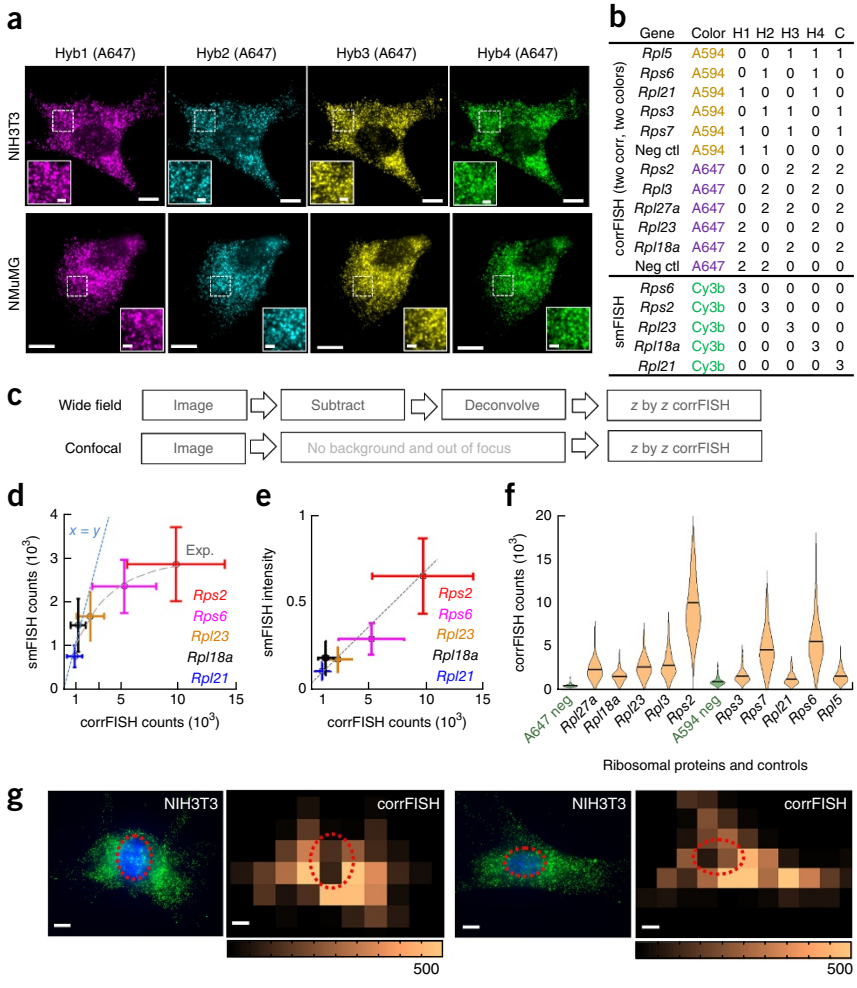

Figure 2 | corrFISH works accurately in cultured cells. (**a**) Ribosomal protein transcripts imaged across four hybridizations in NIH3T3 and NMuMG cells. Scale bars, 10 μm and 2 μm (insets). A647, Alexa-647-labeled probes. (**b**) Ten-gene barcode scheme using two fluorophores and four hybridization rounds (H1–H4). 1, Alexa-594-labeled probes; 2, Alexa-647-labeled probes; 3, Cy3b-labeled probes used one gene at a time for smFISH (localization-based) counting. Neg ctl, negative control. C, positive control, repeat hybridization with hyb3 probe set. (**c**) corrFISH workflows for wide-field and confocal microscopes. (**d**) Comparison of smFISH and corrFISH quantification for *Rpl21*, *Rpl18a*, *Rpl23*, *Rps6*, and *Rps2* demonstrates high dynamic range of corrFISH. A linear $x = y$ line (dashed blue) and an exponential fit (Exp., gray dash) are shown for comparison. n = 85 cells. (**e**) Summed intensity of smFISH dots agrees well with corrFISH quantification (R^2 = 0.95), n = 85 cells. (**f**) Single-cell RNA distributions of ten ribosomal protein genes with negative controls (neg). (**g**) Spatial analysis using corrFISH subregion analysis. Two NIH3T3 cells undergoing *Rpl3* and *Rpl27a* hybridization (round 2) in green; nuclear DAPI, blue. To the right of each cell, transcript abundance map of 25 subregions within the cell. Dashed red circles, nuclear regions. Scale bars, 10 μm.

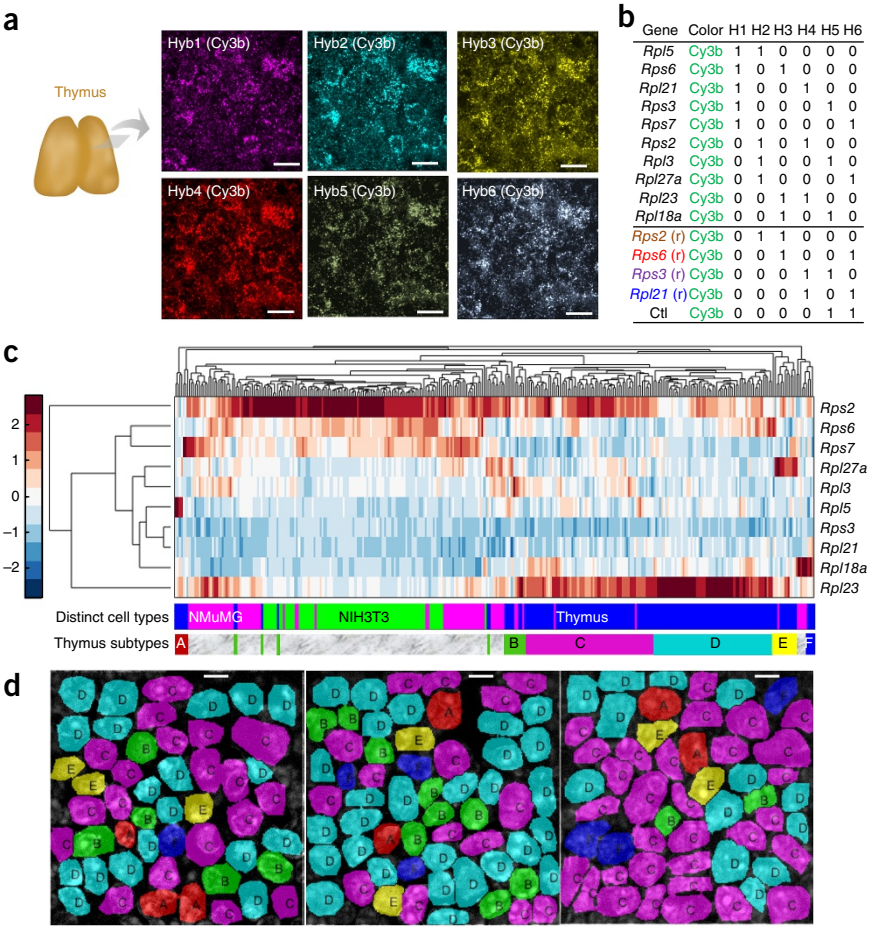

Figure 3 | corrFISH reveals cell-specific ribosomal protein gene expression in tissue sections. (**a**) CorrFISH was applied to mouse thymus sections over six hybridizations using only Cy3b-labeled probes. Scale bars, 10 µm. (**b**) Barcodes for ten ribosomal protein genes, four repeats as a positive control (*Rps2*, *Rps6*, *Rps3*, and *Rpl21*), and empty as a negative control (Ctl). (r), repeat of the same gene. H_1–H_4, hybridization rounds. (**c**) Clustered single-cell gene expression profiles of ten ribosomal proteins from all cell-culture and thymus-tissue-section data (NMuMG, 80 cells; NIH3T3, 85 cells; and thymus, 162 cells). Thymus clusters into six distinct groups of cells (A–F). (**d**) Mapping of the cell clusters in the thymus section at three different locations showed heterogeneous distribution and layering of cell types. Scale bars, 5 µm.

ribosomal proteins in two phenotypically different cell lines: mouse embryonic fibroblast cells (NIH3T3) and mammary-gland epithelial cells (NMuMG) (**Fig. 2a**). We used wide-field and confocal microscopy for our binary barcoded transcript measurements (**Fig. 2b**,**c** and **Supplementary Figs. 6–8**). To assess the dynamic range of corrFISH, we compared smFISH counting to corrFISH quantification for *Rps2*, *Rps6*, *Rpl23*, *Rpl18a*, and *Rpl21* (**Supplementary Fig. 9**). While smFISH counting tended to underestimate the transcripts on account of the overlapping FISH spots (**Fig. 2d**), the integrated intensity correlated well with corrFISH with an R^2 value of 0.95 (**Fig. 2e**). At the highest copy numbers, 5,000 smFISH counts corresponded to 15,000 corrFISH counts. These smFISH comparisons validated that corrFISH achieves reliable detection over a high dynamic range of up to 15,000 counts per cell. In our experiments, a target RNA molecule (one out of three species) with a density of 0.5–9 molecules/µm^2 can be accurately quantified within a total density of 5–15 molecule/µm^2 per hybridization by corrFISH within less than 10% error (**Supplementary Fig. 10**). To determine the sensitivity of corrFISH, we measured the number of empty barcodes detected (**Fig. 2f**). The negative controls in Hyb1 and Hyb2 showed negligible correlation values and provided a limit of detection (LOD) density of 0.08 molecules/µm^2 (222 ± 18 s.e.m. counts/average 2,674 µm^2 cellular area).

corrFISH can also provide spatial maps of gene expression profiles. Subcellular details can be resolved by performing correlation analysis on subregions within cells (**Supplementary Fig. 11**). We profiled transcript distribution in 25 subcellular regions for *Rpl3* and *Rpl27a* mRNAs (**Fig. 2g**). In most fibroblast cells, ribosomal protein transcripts were detected outside the nuclear region (circled) in the subregion correlation analysis. Varying subregion window size did not significantly affect the results (**Supplementary Fig. 12**).

When RNAs are not Poisson distributed across the cell or sample, corrFISH needs to be used with subregion analysis. A heterogeneous image can be subdivided into regions of relatively homogenous, Poisson-distributed mRNAs, which are analyzed separately and summed (**Supplementary Fig. 13**). We successfully implemented spatial analysis in tissue sections at single-cell resolution. However, corrFISH quantitation can break down if RNA granules containing multiple RNAs exist within a diffraction-limited region. These problematic regions can be cropped out of the image and removed from correlation analysis if they occur.

We quantified gene expression of ten ribosomal proteins with corrFISH in 10-µm-thick thymus tissue sections (**Fig. 3**). The thymus is of particular interest because ribosome function has recently been shown to play important roles in hematopoiesis[13] and the immune system[14]. To barcode the ten transcripts, we performed six hybridizations with a single color (**Fig. 3b**). The transcript distributions for these ribosomal protein genes are heterogeneous (**Fig. 3d** and **Supplementary Fig. 14**). Repeat barcodes plotted against their counterparts gave a linear correlation ($R = 0.9$) for 162 thymus cells (**Supplementary Fig. 15**), indicating accurate quantitation in tissues.

To evaluate molecular differences in phenotypically distinct cells, we performed hierarchical clustering of gene expression in 80 NMuMG, 85 NIH3T3, and 162 thymus cells. These results exhibited unique clusters for all three cell types (**Fig. 3c**). The ability to define cell types based on ribosomal protein composition supports the specialized ribosome theory[15]. The NIH3T3 and NMuMG gene expression profiles showed significant variability across single cells and most genes exhibited a unimodal distribution (**Supplementary Figs. 16** and **17**). *Rps2* had the highest

expression level (mean 9,800 copies per cell in NIH3T3 and 5,000 copies per cell in NMuMG), while *Rpl21* and *Rpl18a* had the lowest expression level (respective means of 975 and 1,284 copies per cell in NIH3T3; 981 and 1,400 copies per cell in NMuMG). We observed combinatorial gene expression patterns for each cell type (**Fig. 3c**). For example, *Rps2* (high-copy regulator) and *Rpl21* (low-copy regulator) were relatively uniformly expressed across all cells, while other genes were expressed in a combinatorial set of patterns in different cell types.

To ask whether thymus has spatial patterns of cell types with distinct gene expression profiles, we mapped the transcriptional profiles back to single cells on the tissue images. Unique spatial patterning of six different thymus cell subtypes (A, B, C, D, E, and F) were obtained on three distinct regions of the same thymus section (**Fig. 3d** and **Supplementary Fig. 18**). Cellular interactions in the thymus play an important role in immune cell repertoire formation, and studies of these interactions may be informed by our observation of the spatial organization of cellular subtypes.

To cover a large dynamic range, high-copy-number transcripts can be barcoded and analyzed by corrFISH first. Then the probes can be stripped, and a second barcoding experiment in the same sample can be performed on low-abundance transcripts using localization. This hybrid approach allows low-abundance genes to be accurately detected with single-molecule resolution and high-copy-number genes to be profiled in the same cell.

corrFISH demonstrates that RNA quantification does not need to rely on smFISH spot resolution, removing the constraints on seqFISH to allow the targeting of RNAs with much higher expression levels without the need for super-resolution microscopy. Conceptually, this is similar to the ability of fluorescence correlation spectroscopy (FCS)[16] to quantify molecular concentration down to the single-molecule level even in high-concentration solutions. In our case, the temporal dimension is generated by sequential rounds of hybridization rather than real time. In principle, this image correlation approach can be applied to decode high-density images to multiplex molecular species other than RNA, including proteins[17–20] and metabolites[21].

METHODS

Methods and any associated references are available in the online version of the paper.

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

ACKNOWLEDGMENTS

We thank J. Linton from the Elowitz laboratory (Caltech) for providing cell lines and M. Yui from the Rothenberg Laboratory (Caltech) for the intact thymus organ. We appreciate the help of the City of Hope Pathology Core to slice thymus into sections. This work is funded by US National Institute of Health single-cell analysis program award R01HD075605. A.F.C. is supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

AUTHOR CONTRIBUTIONS

A.F.C. and L.C. designed the project and wrote the manuscript. L.C. supervised the project.

COMPETING FINANCIAL INTERESTS

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

1. Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M. & Cai, L. *Nat. Methods* **11**, 360–361 (2014).
2. Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S. & Zhuang, X. *Science* **348**, aaa6090 (2015).
3. Lee, J.H. *et al. Science* **343**, 1360–1363 (2014).
4. Ke, R. *et al. Nat. Methods* **10**, 857–860 (2013).
5. Lubeck, E. & Cai, L. *Nat. Methods* **9**, 743–748 (2012).
6. Holden, S.J., Uphoff, S. & Kapanidis, A.N. *Nat. Methods* **8**, 279–280 (2011).
7. Zhu, L., Zhang, W., Elnatan, D. & Huang, B. *Nat. Methods* **9**, 721–723 (2012).
8. Schwille, P. *Cell Biochem. Biophys.* **34**, 383–408 (2001).
9. Kettling, U., Koltermann, A., Schwille, P. & Eigen, M. *Proc. Natl. Acad. Sci. USA* **95**, 1416–1420 (1998).
10. Costantino, S., Comeau, J.W.D., Kolin, D.L. & Wiseman, P.W. *Biophys. J.* **89**, 1251–1260 (2005).
11. Kondrashov, N. *et al. Cell* **145**, 383–397 (2011).
12. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. & Tyagi, S. *Nat. Methods* **5**, 877–879 (2008).
13. Signer, R.A., Magee, J.A., Salic, A. & Morrison, S.J. *Nature* **509**, 49–54 (2014).
14. Ito, Y. *et al. Science* **346**, 363–368 (2014).
15. Xue, S. & Barna, M. *Nat. Rev. Mol. Cell Biol.* **13**, 355–369 (2012).
16. Magde, D., Elson, E. & Webb, W.W. *Phys. Rev. Lett.* **29**, 705–708 (1972).
17. Gerdes, M.J. *et al. Proc. Natl. Acad. Sci. USA* **110**, 11982–11987 (2013).
18. Murray, E. *et al. Cell* **163**, 1500–1514 (2015).
19. Gong, H. *et al. Bioconjug. Chem.* **27**, 217–225 (2016).
20. Fan, R. *et al. Nat. Biotechnol.* **26**, 1373–1378 (2008).
21. Xue, M. *et al. J. Am. Chem. Soc.* **12**, 4066–4069 (2015).

ONLINE METHODS

Theory. For quantification of transcripts, we used correlation analyses on a series of seqFISH images. We performed a pairwise correlation from the subset of hybridization images for quantitation of each gene. For instance, we used equation (1) to compute crosscorrelation in between Hyb1 and Hyb2 images[8–10] (see **Supplementary Note** for derivation).

$$G_{12} = \frac{\mathcal{F}^{-1}\{\mathcal{F}(H_1)\mathcal{F}^{*}(H_2)\}}{\langle H_1\rangle\langle H_2\rangle} - 1 \quad (1)$$

where G_{12} is the crosscorrelation of the images from hyb1 and hyb2, $\mathcal{F}^{-1}$ is the inverse Fourier Transform and $\mathcal{F}$ is for the Fourier Transform operations. H_1 and H_2 are the images, and $\langle H_1\rangle$ and $\langle H_2\rangle$ are the mean values of the corresponding images (brackets are expected values).

These correlation results were then converted to the transcript abundance by using equations (2) and (3). The amplitude of cross-correlation was normalized by the autocorrelation amplitudes as an estimator of RNA copy number in a cell-per-point spread function (PSF) area.

$$\langle d_{12}\rangle = \frac{G_{12}}{G_{11}G_{22}} \quad (2)$$

where $\langle d_{12}\rangle$ is the abundance of transcripts in an PSF area that are common across hybridizations 1 and 2 (brackets are expected values), G_{12} is the amplitude of the crosscorrelations peak between the Hyb1 and Hyb2 images, and G_{11} and G_{22} are the autocorrelation peak amplitudes of Hyb1 and Hyb2, respectively. Transcript count per PSF area was then converted to the total number of transcripts for a single cell:

$$\langle N_{12}\rangle = \frac{\langle d_{12}\rangle}{A_{PSF}} \times A_{\mathrm{Cell}} \quad (3)$$

where $\langle N_{12}\rangle$ is the number of transcripts that are common across hybridizations 1 and 2. Here, $\langle d_{12}\rangle$ is the copy number of transcripts per PSF area from equation (2). A_{PSF} is the PSF area ($\pi \times PSF_{width}{}^2$), where PSF_{width} is about 0.3 µm. A_{Cell} is the area of the single-cell image ($N \times p^2$), where N is the total number of pixels in a cell and p is the pixel size of 0.13 µm in our corrFISH analysis.

Simulations. We used digital data to imitate the realistic FISH experiments at various densities and conditions. Using a custom-written MATLAB algorithm, point emitters were randomly distributed in a digital image of size 230 × 230 pixels (corresponding to 30 µm × 30 µm cell area). Emitters were convolved with 0.3 µm width point-spread function (PSF) corresponding to a 100× wide-field fluorescence microscope (0.13 µm pixel size), matching the cell sizes and pixel resolution in the FISH experiments. The obtained images mimic the molecules labeled with fluorescent dyes. Adjusting the total number of emitters per simulation area changed concentration of the molecules. Sequences of these images were correlated in spatial or Fourier domains to compute the abundances of molecules that are common across these image arrays (**Supplementary Note**). In addition, we performed super-resolution microscopy simulations with 0.05 µm pixel size to show that correlation decoding can even achieve higher-density multiplexing (**Supplementary Fig. 19**).

Cultures. NIH3T3 cells (ATCC CRL-1658) were cultured in Dulbecco's Modified Eagle's medium (DMEM) with GlutaMAX, high glucose, and pyruvate (Thermo Fisher Scientific, Gibco, 10569), bovine calf serum (10%), and penicillin–streptomycin (PenStrep) media and passaged every few days. Cells were plated on 24 × 50 mm fibronectin coated slides overnight within culture media in a petri dish. After 15 h, glass slides were washed with phosphate-buffered saline (PBS) 1× and then fixed with 4% formaldehyde for 4 min. Cells were then placed in 70% ethanol for permeabilization and stored at −20 °C. NMuMg cells (ATCC CRL-1636) were cultured in DMEM with glutamine and 4.5 g/L glucose (Thermo Fisher Scientific, Gibco, 11965), fetal bovine serum (10%), insulin (10 mcg/ml), and PenStrep media. The rest of the fixation and permeabilization protocols was the same. For experiments, glass slides with either cell type were removed from −20 °C storage and then dried with air blow. A custom flow-cell chamber (Grace Bio-Labs, SecureSeal, 3 mm × 11 mm size × 0.5 mm depth, RD478682) was then bound to this glass slide for sequential FISH measurements.

Tissue. Intact thymus was extracted from a four-week-old female mouse. Animal handling was done in Rothenberg's laboratory with approval by Caltech's Institutional Animal Care and Use Committee. Immediately after extraction, thymus was embedded into 4% paraformaldehyde for three hours at room temperature for fixation of cells within the organ. Thymus was then rinsed with PBS1× for 10 min. For slicing, thymus was then embedded in 10% sucrose and kept at 4 °C for more than 12 h. Saline coated glass slides (24 mm × 50 mm) were then used to host 10 µm tissue sections after cutting at the City of Hope Pathology Core. Thymus tissue sections were kept at −80 °C deep freezer for storage. The day before the FISH experiment, tissue section was removed from the freezer and thawed at room temperature for one hour. To clear the section and reduce nonspecific binding, tissue was first treated with 8% sodium dodecyl sulfate (SDS) for 10 min. Tissue section was then treated with 70% ethanol overnight at 4 °C for permeabilization. The next day, the glass slide with tissue section was then bound to the hybridization flow chamber (Grace Bio-Labs, SecureSeal, 8–9 mm diameter × 0.6 mm depth, GBL621101) for sequential labeling and imaging experiments.

SeqFISH. 24 FISH probes were used for each gene. Cy3B, Alexa 594, and Alexa 647 were coupled to the probes according to the barcode scheme. Cells or tissue sections within flow chambers were hybridized at a concentration of 1 nM per probe overnight in a hybridization buffer of 10% dextran sulfate (Sigma D8906), 30% formamide, 2× saline-sodium citrate (SSC) buffer at room temperature. Labeled cells were then washed with 30% formamide for 30 min. For tissues, cells were washed a second time with 50% formamide for 10 min. Samples were rinsed several times with 2× SSC to remove leftover probes. Samples were then imaged within an antibleaching buffer consisting of Tris-HCL (20 mM), NaCl (50 mM), glucose (0.8%), saturated Trolox (Sigma, 53188-07-1), pyranose oxidase (Sigma: P4234, OD405 nm: 0.05), and catalase (Sigma, 9001-05-2, 1:1000 dilution). A piece of glass slide or parafilm was used to cover the top facet of the flow cell, preventing the

imaging or hybridization buffer from evaporation. After imaging the samples, FISH probes were removed by 100 units of DNase I enzymatic treatment (Sigma-Aldrich, 04716728001 ROCHE) for 4 h followed by a postwash with 30% formamide and 2× SSC wash for 10 min. The cells were subsequently hybridized by another probe set at 1 nM concentration for more than 12 h at room temperate in the hybridization buffer. These presented imaging, probe stripping, and hybridization protocols are repeated based on the barcoding scheme[1].

Barcoding. In cultures, we targeted transcripts for ten ribosomal proteins with corrFISH with two color binary codes (**Fig. 2b**). We barcoded five of the ribosomal protein genes (*Rpl5*, *Rps6*, *Rpl21*, *Rps3*, and *Rps7*) by sequential hybridizations of probes labeled by Alexa 594 fluorophores, and then we coded the second set of genes (*Rps2*, *Rpl3*, *Rpl27a*, *Rpl23*, and *Rpl18a*) by Alexa 647 probes. This implementation of corrFISH keeps the barcoding within each fluorophore channel to avoid errors introduced by chromatic aberration and spectral crosstalk. For validation of corrFISH, we used an additional color (in this case Cy3b) for single-molecule FISH experiments to validate the corrFISH results (**Fig. 2b**). In tissues, we barcoded transcripts for ten ribosomal proteins with a single binary code (Cy3b), in which additional four repeats for *Rps2*, *Rps6*, *Rpl21*, and *Rps3* as a positive control, and 1 empty barcode as a negative control are included in the barcoding scheme (**Fig. 3b**). Adding more colors and correlation dimensions to the barcoding scheme linearly increased the number of target RNA species for corrFISH (**Supplementary Fig. 5**). Incorporating chromatic corrections can further scale up the multiplexing capacity to the exponential rate seen in the seqFISH method[1].

Imaging. Cell cultures were imaged by a custom wide-field fluorescence microscope with single-molecule imaging capability (**Supplementary Fig. 20**). A simple fiber combiner (CNI Laser, 7 pieces to 1 piece connect with ferrule connector (FC) end or, alternatively, Oz Optics, 1 × 6 pigtail wavelength division multiplexer) was used to merge laser beams creating a single output beam for illumination. Each illumination laser was fiber coupled and had about 1 W power with 405, 473, 533, 589, 640, and 730 nm center wavelengths (CNI Lasers). This fiber-based design removed the free space optics alignment requirement for multicolor illumination of cells. Fiber output (>100 mW) was then collimated with a simple convex lens and got vibrated by a custom fiber shaker all the time to avoid speckle issues. To enlarge the beam, a 1.5× telescope (f_1: 200 mm and f_2: 300 mm) was inserted after collimation lens. Expanded illumination was then focused at the back aperture of the objective lens using another 200 mm convex lens. All of these lenses (Thorlabs) were 2″ diameter to reduce aberrations. 60× objective (1.42 numerical aperture) lens was used to collect fluorescence from the cells. We obtained 96× magnification with the use of 1.6× tube lens magnification providing 0.135 µm pixel size. High-quantum-efficiency Andor Ikon M camera was used to record the fluorescence from single molecules. We used up to 500 ms exposure time per color channel to capture FISH images. All the devices were controlled by Micromanager.

Tissue sections were imaged by a spinning disk confocal microscope (Andor, WD model). In our experiments, 561 nm laser illumination was used to image Cy3b-labeled probes and 405 nm laser was used to capture DAPI signal. 100× objective lens was used to collect the fluorescence. We tuned into relatively higher exposure time of up to 20 s per FISH image to obtain high signal-to-noise ratio FISH dots. A Metamorph software was used to control the devices and capture images. Compared to the point confocal imaging, spinning disk allowed us to reduce the photobleaching of our single molecules and increased imaging speed.

Analysis. Sequential FISH images were processed in MATLAB using a digital image-processing pipeline. Raw single-molecule images were segmented to create a mask to cover the entire cell. Cell autofluorescence background was subtracted by using the image of the cell without FISH probes. Cell background image was registered to the original cell image with FISH probes to improve the subtraction quality. To remove the out-of-focus light, background subtracted images were deconvolved using a Lucy–Richardson iterative algorithm. The binary mask was then multiplied by deconvolved cell images from sequential hybridizations to define the cell areas. corrFISH was then processed *z* by *z* plane on the cell images. These transcript counts were then combined by summing the results from each section (**Supplementary Fig. 8**). Here, two subsequent planes were summed to increase the robustness of corrFISH processing. The final single-cell transcript counts were then plotted as a heat map, violin plot, and hierarchical cluster of gene expression. Additional control experiments with various experimental conditions were analyzed by box plots (**Supplementary Figs. 21** and **22**).

Single-molecule FISH counting was performed based on a localization and counting algorithm[12]. Specifically, a Laplacian of Gaussian (LoG) filter was applied to the background subtracted and deconvolved images (**Supplementary Note**). Individual transcripts were then identified and counted *z* by *z* plane. These counts were then summed to compute the gene expression in single cells. As the density of molecules increased, the FISH dots overlapped, creating underestimation for most of the ribosomal protein transcripts.

For spatial analysis in cultured cells, repeat hybridizations, both containing *Rpl3* and *Rpl27a*, are correlated to generate the map (**Fig. 2g**).

Tissue analysis was performed by segmenting the cells based on their DAPI image and propagating to subsequent hybridizations. We processed 16 pixel by 16 pixel subregions and the subregions within a cell mask were correlated individually. Then the correlation values were summed together to provide the gene expression value for that cell (**Supplementary Fig. 13**).

For statistical analysis, we used Origin and Excel programs with regression and Student's *t*-test modules.

Code availability. Custom MATLAB source codes (**Supplementary Software**) with test images are available and updated at https://github.com/singlecelllab/correlationFISH.

doi:10.1038/nmeth.3895