Supplementary Materials for

Dense transcript profiling in single cells by image correlation decoding

Ahmet F. Coskun[1,2] and Long Cai[1,2]

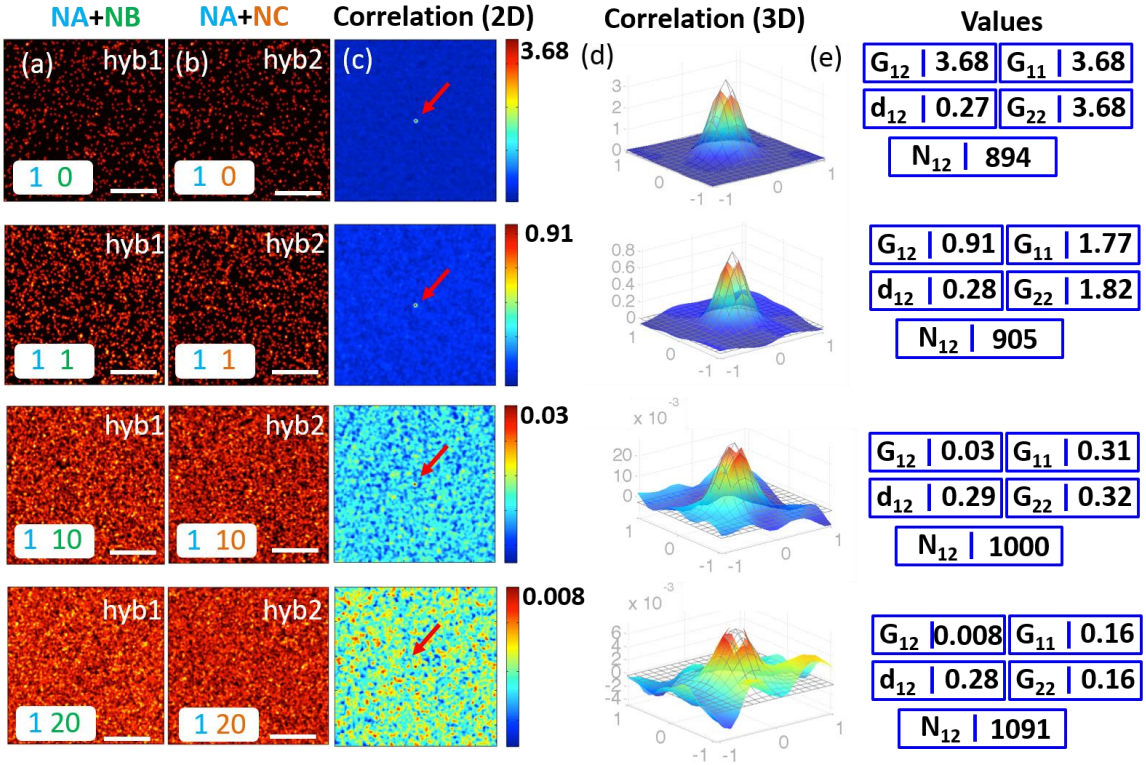
[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA.

[2]Program in Biochemistry and Molecular Biophysics, California Institute of Technology, Pasadena, CA.

Supplementary File	Title
Supplementary Figure 1	Simulated FISH images at various transcript densities and the corresponding correlation results
Supplementary Figure 2	Comparison of correlation and localization approaches for dense transcript quantitation using simulations
Supplementary Figure 3	Coefficient of variation analysis for detection of 1% barcoded transcript within simulated data
Supplementary Figure 4	Five gene to five gene experiments, one gene shared, to validate high density condition of corrFISH
Supplementary Figure 5	Barcoding capacity analysis
Supplementary Figure 6	Auto-fluorescence background subtraction
Supplementary Figure 7	Robust custom developed z by z plane analysis
Supplementary Figure 8	CorrFISH digital processing work-flow for two different imaging schemes of wide-field and confocal imaging
Supplementary Figure 9	Validation of corrFISH by smFISH counting and intensity analysis for five ribosomal protein genes
Supplementary Figure 10	Cell area and transcript density histograms in cultures
Supplementary Figure 11	Spatial analysis strategy using subregions
Supplementary Figure 12	Subregion windows size effect on corrFISH results
Supplementary Figure 13	corrFISH quantification in tissues by spatial analysis
Supplementary Figure 14	corrFISH in multiple layers of tissue section and transcript distributions
Supplementary Figure 15	Repeat hyb analysis in corrFISH analysis in thymus
Supplementary Figure 16	RNA copy number distributions and clustering in NMuMG and NIH3T3
Supplementary Figure 17	RNA density distributions in NMuMG and NIH3T3 cells
Supplementary Figure 18	Spatial transcript distribution maps in three locations of a thymus tissue section.
Supplementary Figure 19	Image correlation analyses to improve the density of transcript detection in super resolution microscopy
Supplementary Figure 20	Experimental setup for a simplified single molecule-imaging microscope using a fiber combiner and lasers
Supplementary Figure 21	Detection of *Rps2* gene within total increasing transcript density
Supplementary Figure 22	Detection of *Rps2* gene at different exposure times of 50 ms, 100 ms, and 200 ms
Supplementary Methods	
Supplementary Note	

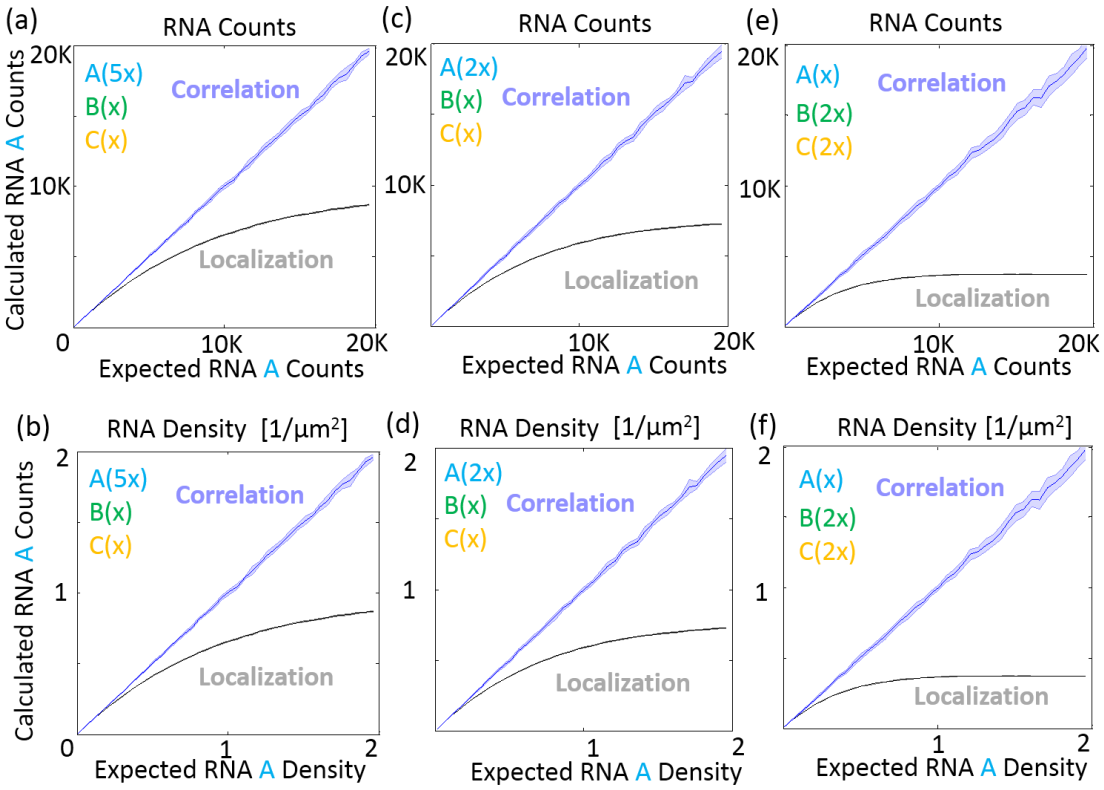
Nature Methods: doi:10.1038/nmeth.3895

Simulated FISH images at various transcript densities and the corresponding correlation results



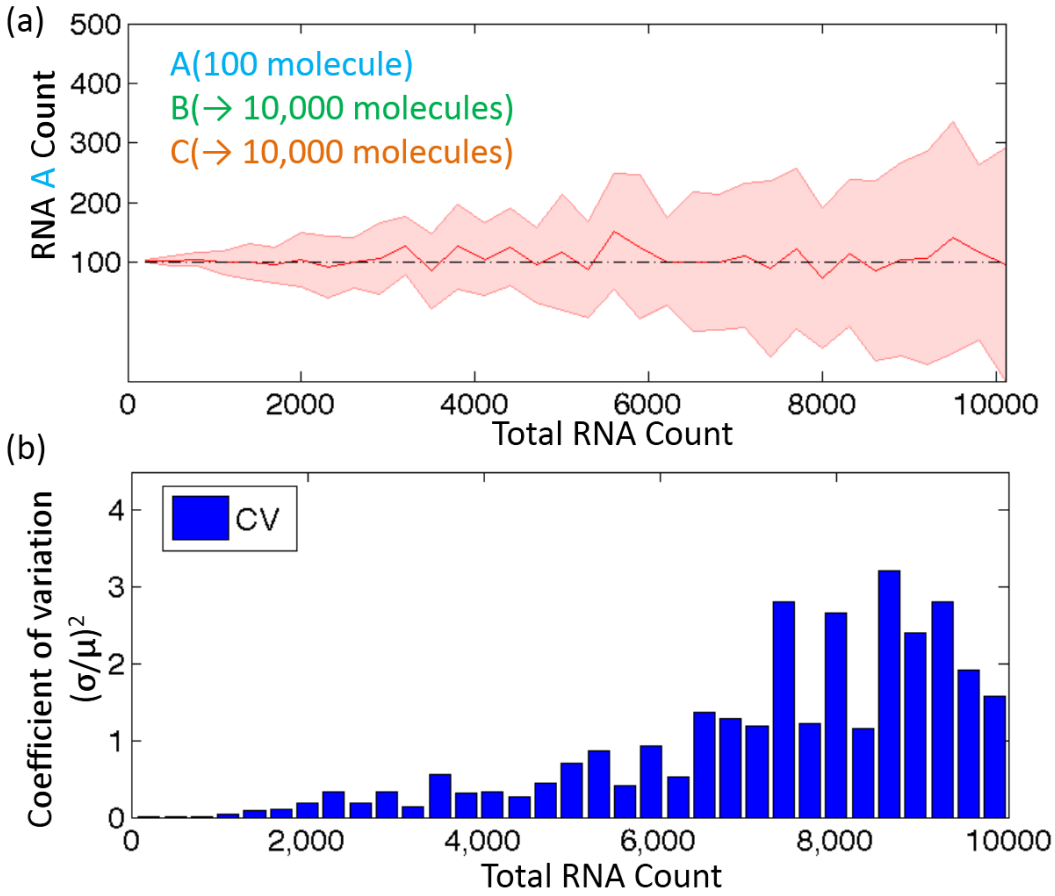
Supplementary Figure 1. Simulated FISH images in 30 µm area and their correlations. (a) Hyb1 images include transcripts A and B. Density of RNA A was fixed at 1 molecule while B was changed up to 20 molecule μm^{-2}. (b) Hyb2 images cover same A transcripts but different sets of C transcripts compared to hyb1 ones. (c) The correlation of hyb1 and hyb2 images provides a correlation matrix with a peak value in the center of the image (denoted by red arrow). Rest of the image is the background corresponding to the noise in the correlation function. Correlation noise scales up with the square root of the number of uncorrelated species. The detection of transcripts is limited by the noise competing with the peak value of the correlations. (d) A Gaussian function was then fit on to the 2 µm x 2 µm central region of the 2D correlation matrix. Gaussian fit results are presented in 3D with raw image in color and overlaid with gray best fit lines. (e) Corresponding correlation amplitude values (G_{11}, G_{22}, G_{12}), molecule densities (d_{12}) and molecule counts (N_{12}).

Comparison of correlation and localization approaches for dense transcript quantitation using simulations



Supplementary Figure 2. Performance comparison of correlation and localization processing methods at various transcript densities. Correlation achieves much more accurate estimations of RNA levels even at higher density in comparisons to localization methods. Up to 20,000 counts or 2 molecule per μm^2 density shared A transcripts were measured with both approaches in simulated images in the presence of up to uncorrelated RNA B and C species in two subsequent hybridization. The ratio of the A and B/C was varied from (a-b) 5:1, (c-d) 2:1, and (e-f) 1:2. Increasing relative amount of uncorrelated species compared to the correlated ones in images caused the localization method to fail in lower densities while the image correlation strategy was not significantly affected.

Coefficient of variation analysis for detection of 1% barcoded transcript within simulated data



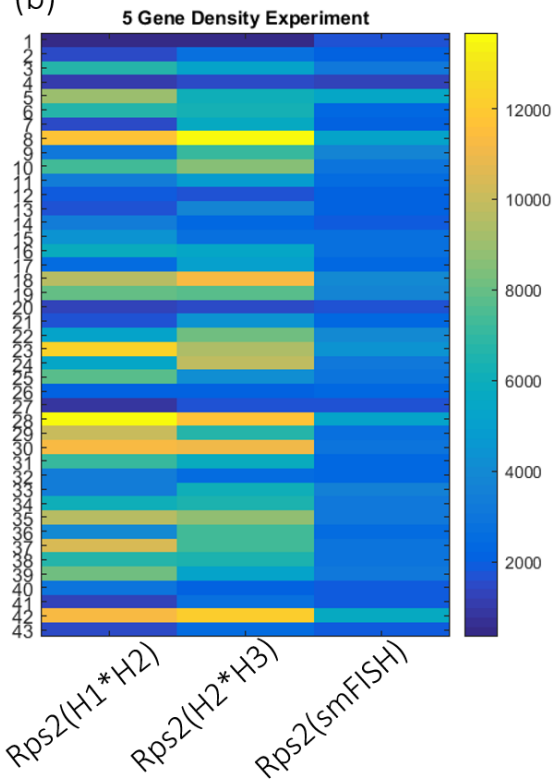
Supplementary Figure 3. Transcript detection within total RNA density in 30 µm simulation cell area. RNA A was kept constant at 100 molecules while RNA B and C concentration was varied up to 10,000 molecules in hyb 1 and hyb2 images, respectively (*n*: 20 simulation runs at each total RNA count values). (a) The detected mean value for 100 molecules as the total transcript counts increase. (b) The coefficient of variation (CV) analysis for detection of 100 molecules. The bootstrapped *CV* for 100 A in 10,000 B or C was 1.76±1.54 (s.e., *n*=20).

Five gene to five gene experiments, one gene shared, to validate high density condition of corrFISH

(a)

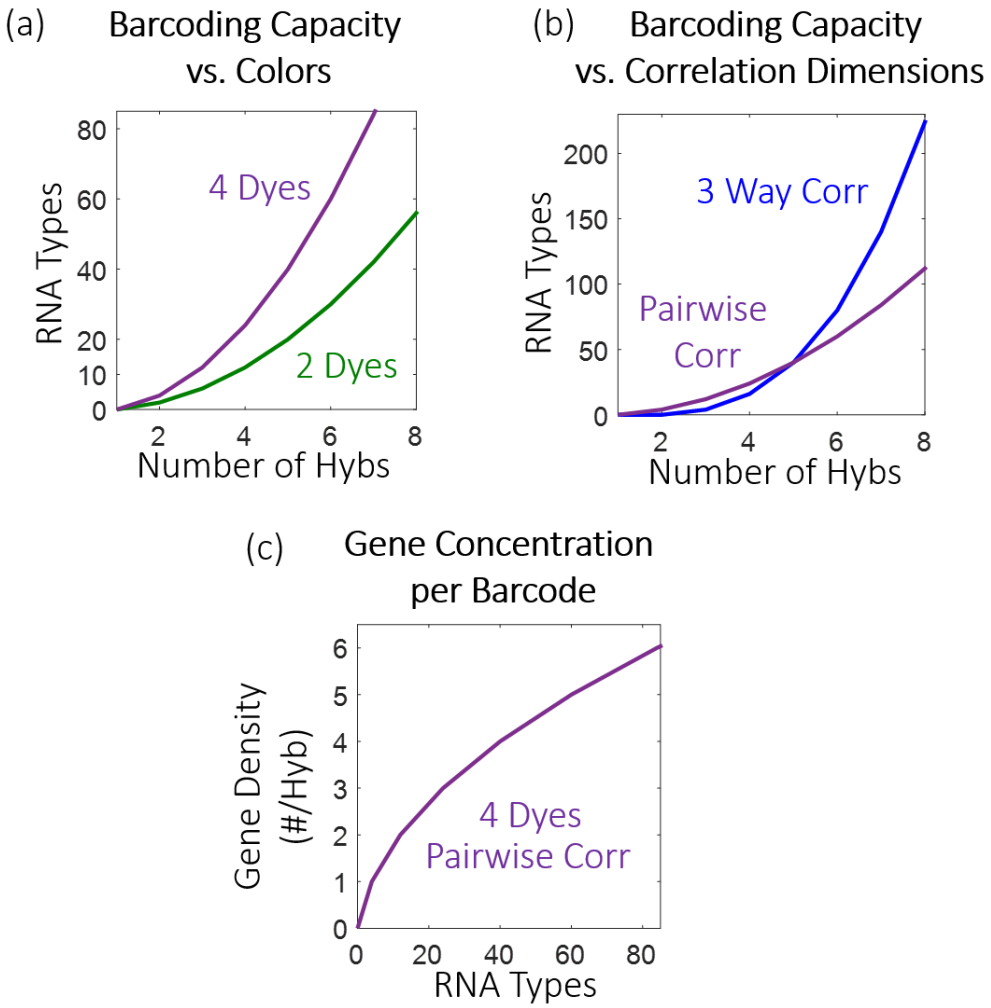
Gene	Color	H1	H2	H3
Rps2	A647	1	1	1
Rps7	A647	1	0	0
Rpl27a	A647	1	0	0
Rpl3	A647	1	0	0
Rpl21	A647	1	0	0
Rps3	A647	0	1	0
Rpl18a	A647	0	1	0
Rpl23	A647	0	1	0
Rpsa	A647	0	1	0

(b)



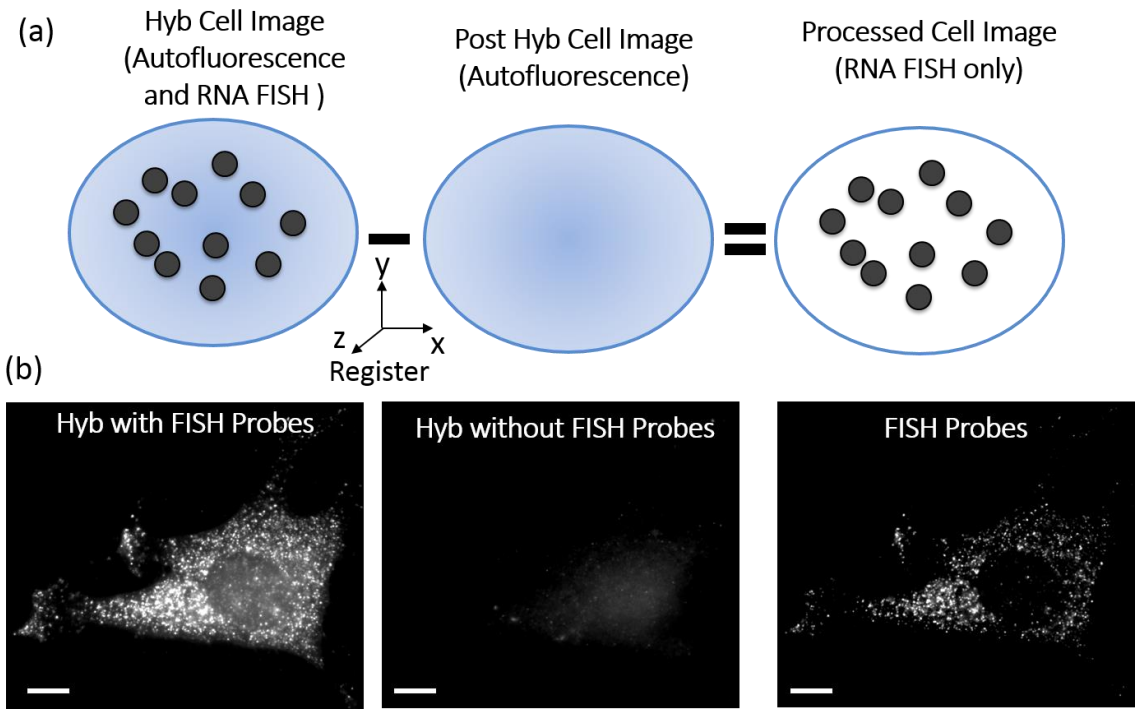
Supplementary Figure 4. A coding experiment using five gene per hybridization to mimic high density needed for multiplexing a large number of genes with the corrFISH barcoding scheme. (a) Two hybs were barcoded with five gene per hybridization and only one gene Rps2 is shared. The third hyb included *Rps2* only. (b) Heat map for gene expression of NIH3T3 cells. Rps2 is calculated from hyb1 and hyb2 (H1*H2), from hyb2 and hyb3 (H2*Rps2 from H3), and from smFISH counting of Hyb3. *n*=43 cells.

Barcoding capacity analysis



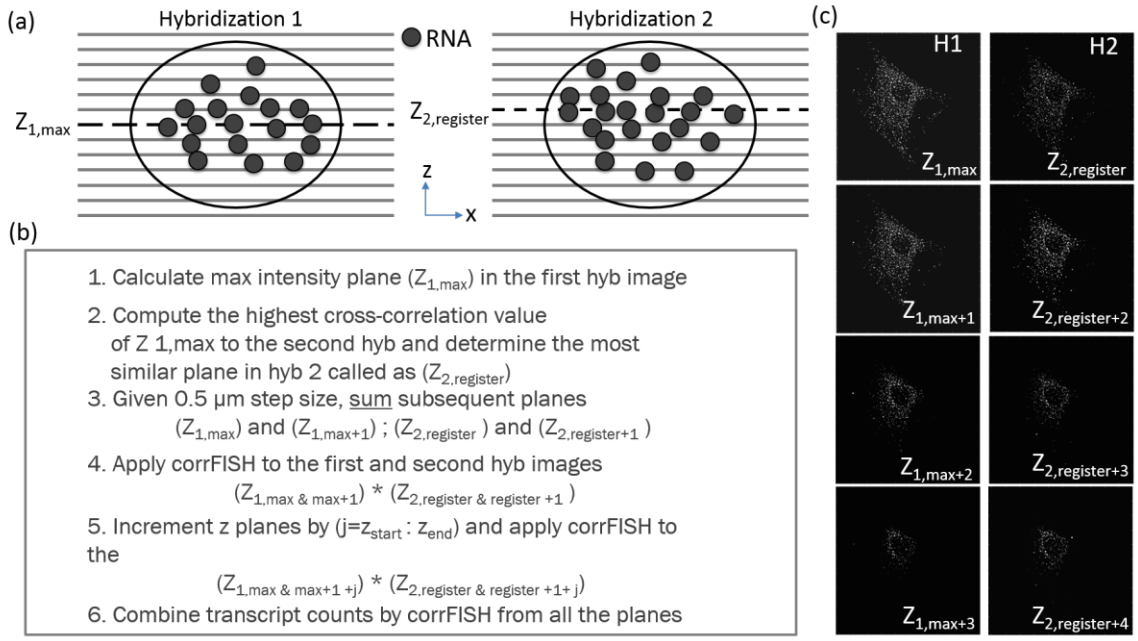
Supplementary Figure 5. corrFISH barcoding capacity scales with number of dyes and correlation dimensions. (a) Increasing colors and (b) increasing correlations among more hybridizations can barcode additional RNA species. (c) Using four colors and pairwise correlations, corrFISH can cover 80 genes by labeling only six genes per hybridization. Thus, the density of transcripts per hybridization is similar to the five genes per hybridization demonstrated in this paper.

Auto-fluorescence background subtraction



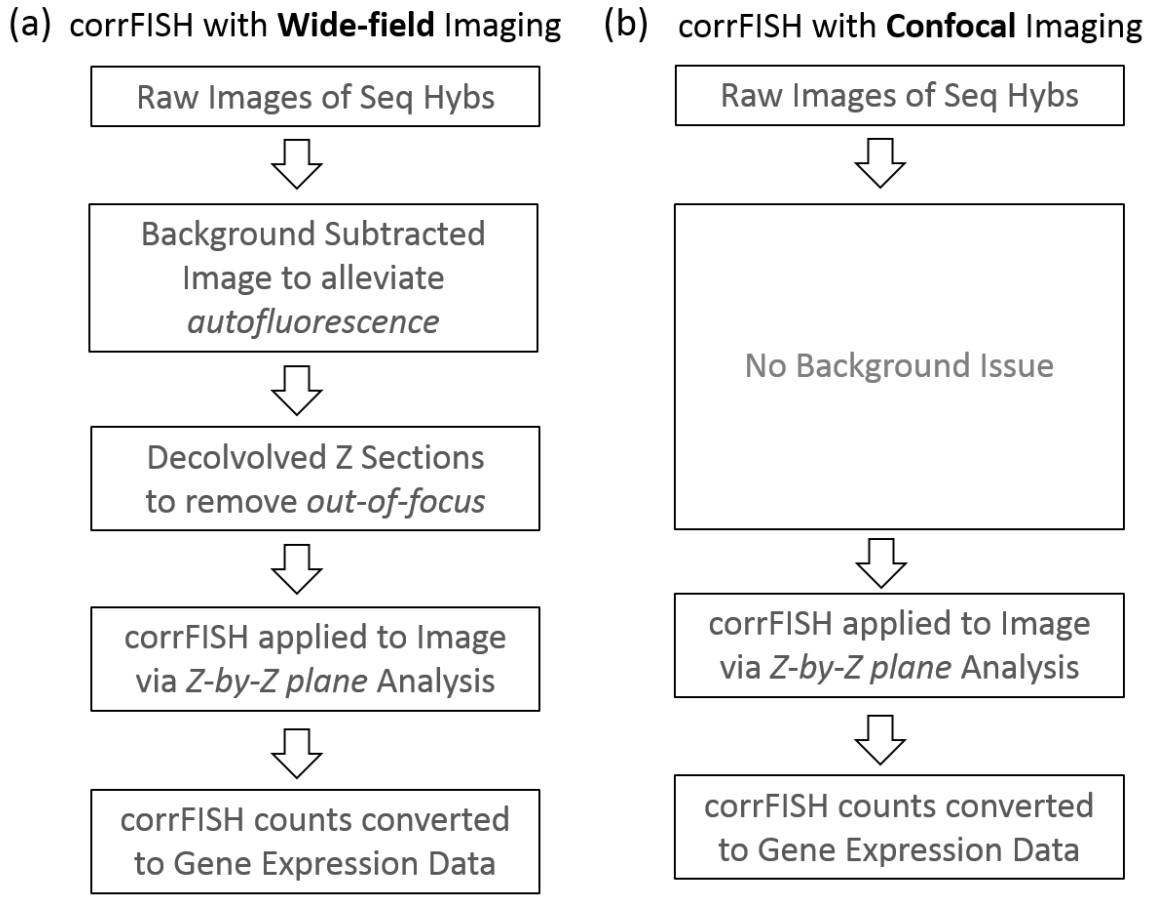
Supplementary Figure 6. Auto-fluorescence background subtraction strategy. (a) Schematic demonstration of a cell image with autofluorescence signal and FISH dots, which was subtracted from cell image without FISH probes. The resultant cell image only includes contribution from RNA FISH signal. (b) Experimental data in NIH3T3 cells using Alexa 594 dyes for repeat control hyb3 in ten gene measurements. The cell image with FISH probes and background signal due to the autofluorescence, which was similarly subtracted from the same cell image with autofluorescent background only. The final image includes only FISH dots without background signal. Scale bar is 10 µm. Note that all the experimental images were contrasted the same.

Robust custom developed z by z plane analysis



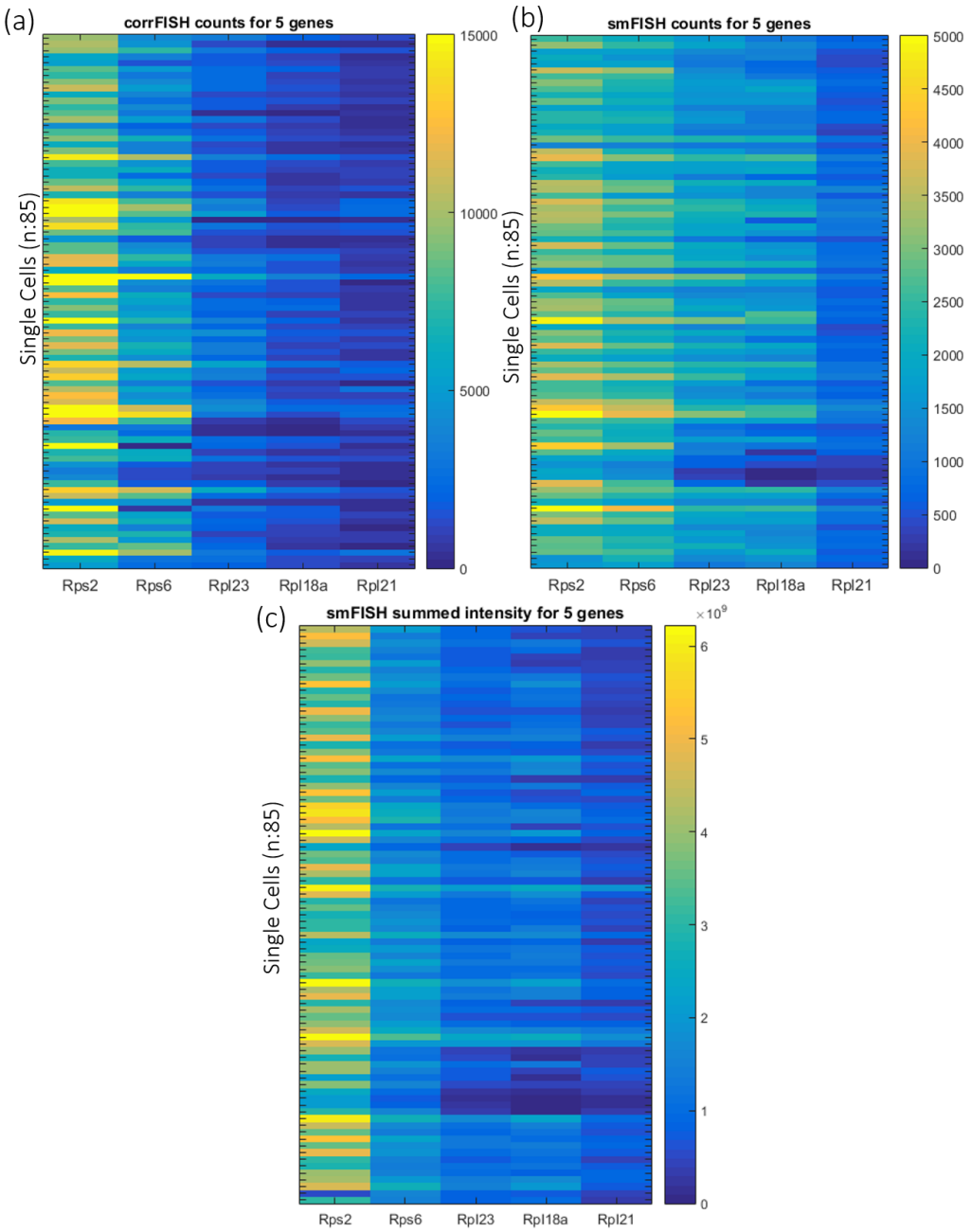
Supplementary Figure 7. A custom developed corrFISH Z-by-Z plane analysis. (a) X-Z section of a cell for hybridizations 1 and 2. Maximum intensity Z section was determined from the first hyb ($Z_{1,max}$) and then registered to the most similar intensity distribution plane in the second hyb ($Z_{2,\ register}$) for corrFISH analysis. (b) Step by step processing for Z-by-Z corrFISH analysis. After defining $Z_{1,max}$, $Z_{2,\ register}$ was computed using highest value of normalized cross-correlation amplitude. Two subsequent planes were summed to create a biplane image per hybridization. corrFISH was processed on these biplanes. Using binary increments on z planes, corrFISH was scanned over Z sections from starting Z section to the ending Z section. Finally, corrFISH counts for biplanes were summed to calculate the transcript abundance per cell. (c) An NIH3T3 cell image was shown for four different subsequent biplanes.

CorrFISH digital processing work-flow for two different imaging schemes of wide-field and confocal imaging



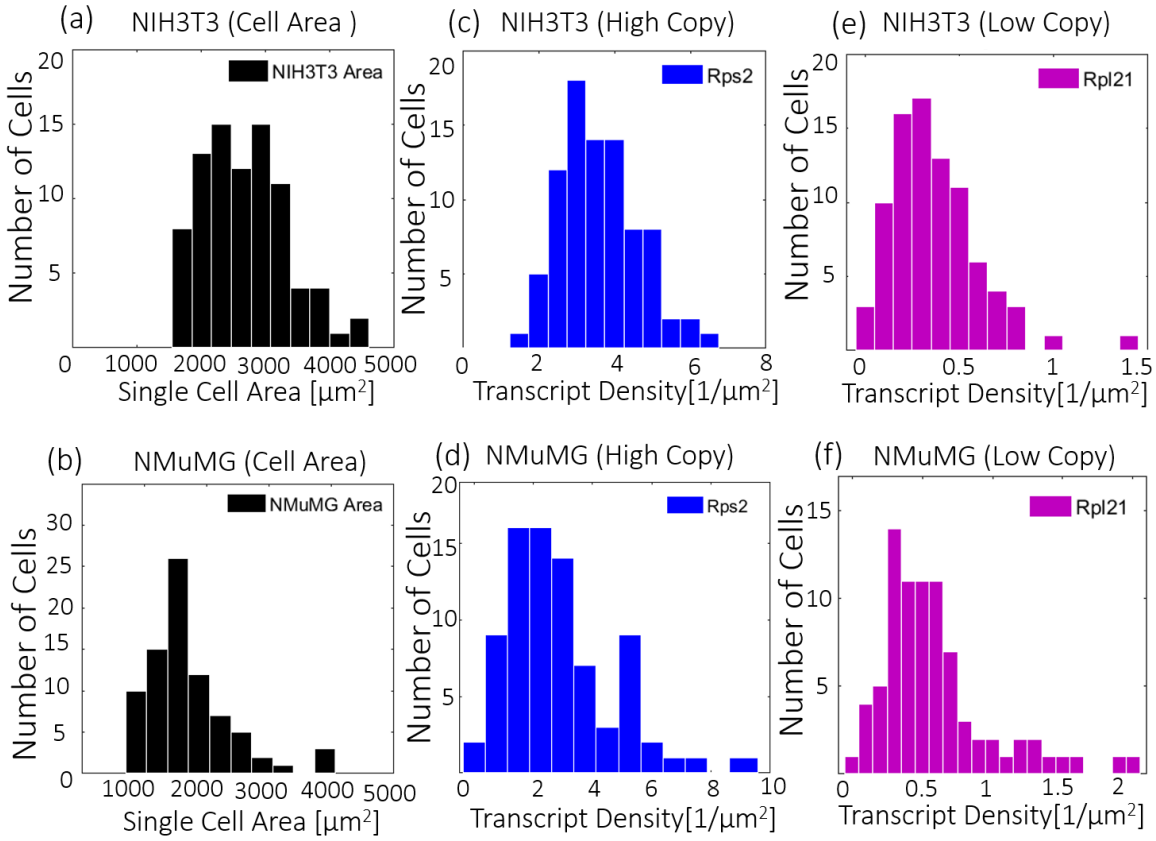
Supplementary Figure 8. Image analysis work-flow of corrFISH for two different imaging modalities. (a) For wide field imaging, first raw cell images were subtracted from the cell background without FISH probes to remove autofluorescence (**Supplementary Figure 6**). The background subtracted cell image was then deconvolved at each z section to remove out-of-focus light. corrFISH was then processed using custom developed z-by-z analysis (**Supplementary Figure 7**). corrFISH counts were summed to compute gene expression per cell. (b) For confocal imaging, since there is no out of focus light, corrFISH was directly applied to the image via Z-by-Z plane analysis for gene expression analysis.

Validation of corrFISH by smFISH counting and intensity analysis for five ribosomal protein genes



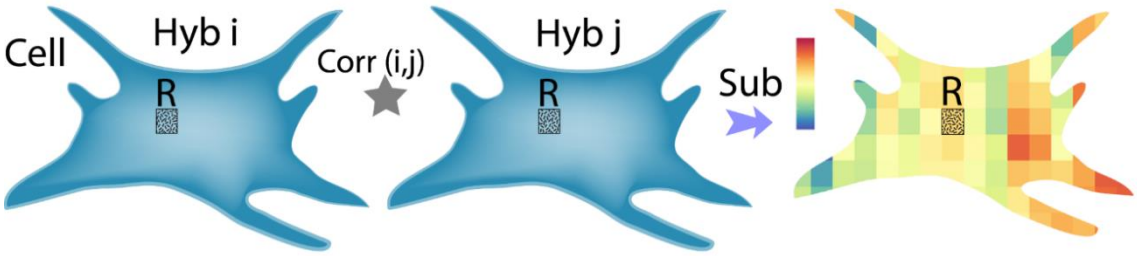
Supplementary Figure 9. Validation of corrFISH by comparison to smFISH in NIH3T3 cultures (*n*=85 cells), the plot of which is presented in **Figures 2d and 2e**. Corresponding heat maps for five different ribosomal proteins: *Rps2*, *Rps6*, *Rpl23*, *Rpl18*, and *Rpl21*. Highest to lowest gene expression patterns for single cells agree well with each other in all the cases. (a-b) The most abundant smFISH counts of 5,000 corresponded to 15,000 corrFISH abundances, suggesting 3-fold difference in between the two due to undercounting in smFISH. (a-c)The smFISH summed intensity providing a linear correlation to the corrFISH counts.

Cell area and transcript density histograms in cultures

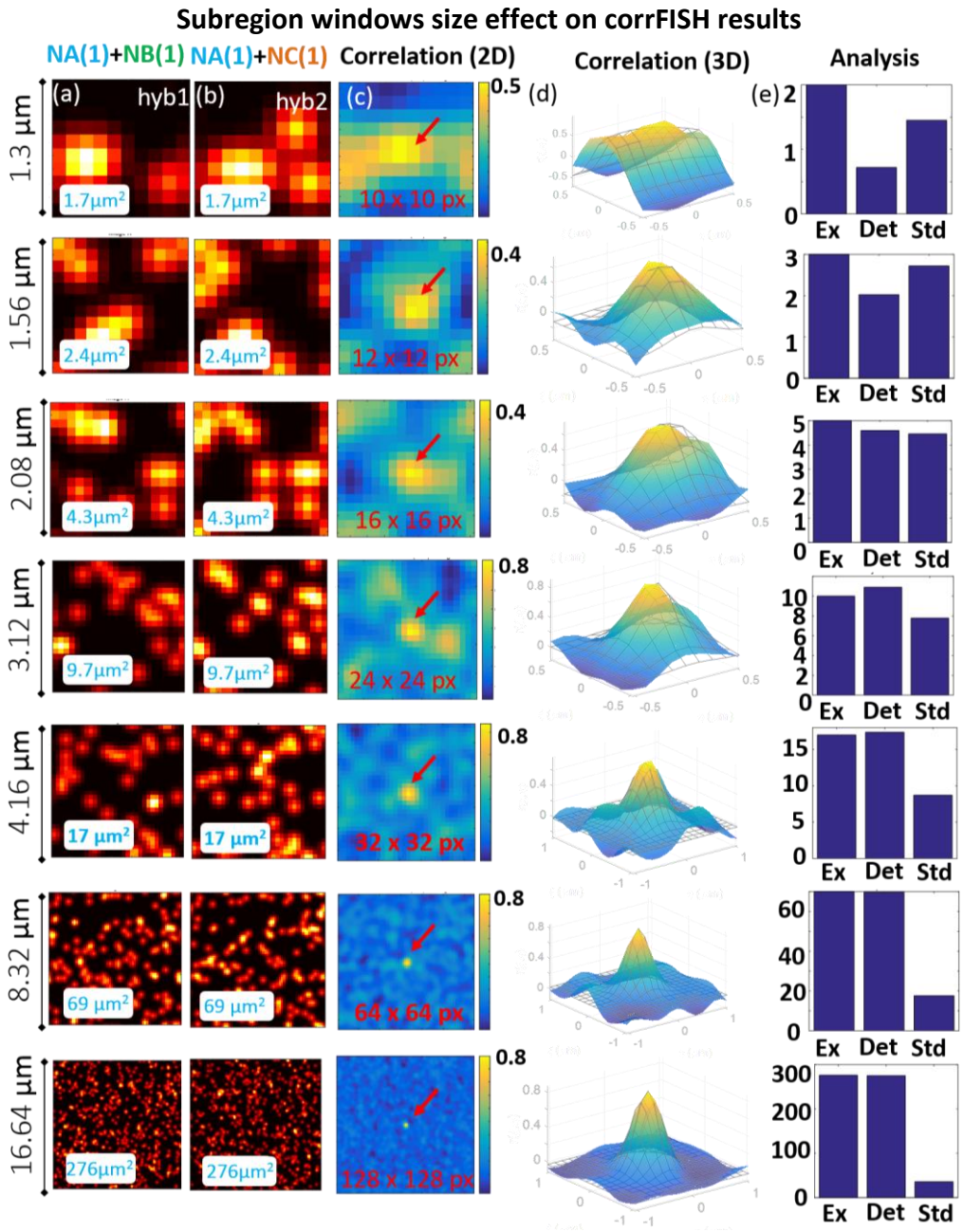


Supplementary Figure 10. Experimental transcript density measurements of ribosomal protein genes in NIH3T3 and NMuMG cells. (a-b) Histogram of total cell areas (*n*=85 cells) ranging from 1,500 to 5,000 μm^2 per NIH3T3 cell, and 1,000 to 4,000 μm^2 per NMuMG cell. (c-d) Transcript density (transcript copy number/cell area) histograms of high copy gene (*Rps2*) varying from 0.5 to 7 μm^{-2} and 0.5 to 9 μm^{-2} per cell, respectively. (c) Corresponding density for low copy gene (*Rpl21*) varying from 0.02 to 1.5 μm^{-2} and 0.02 to 2 μm^{-2} per cell, respectively.

Spatial analysis strategy using subregions

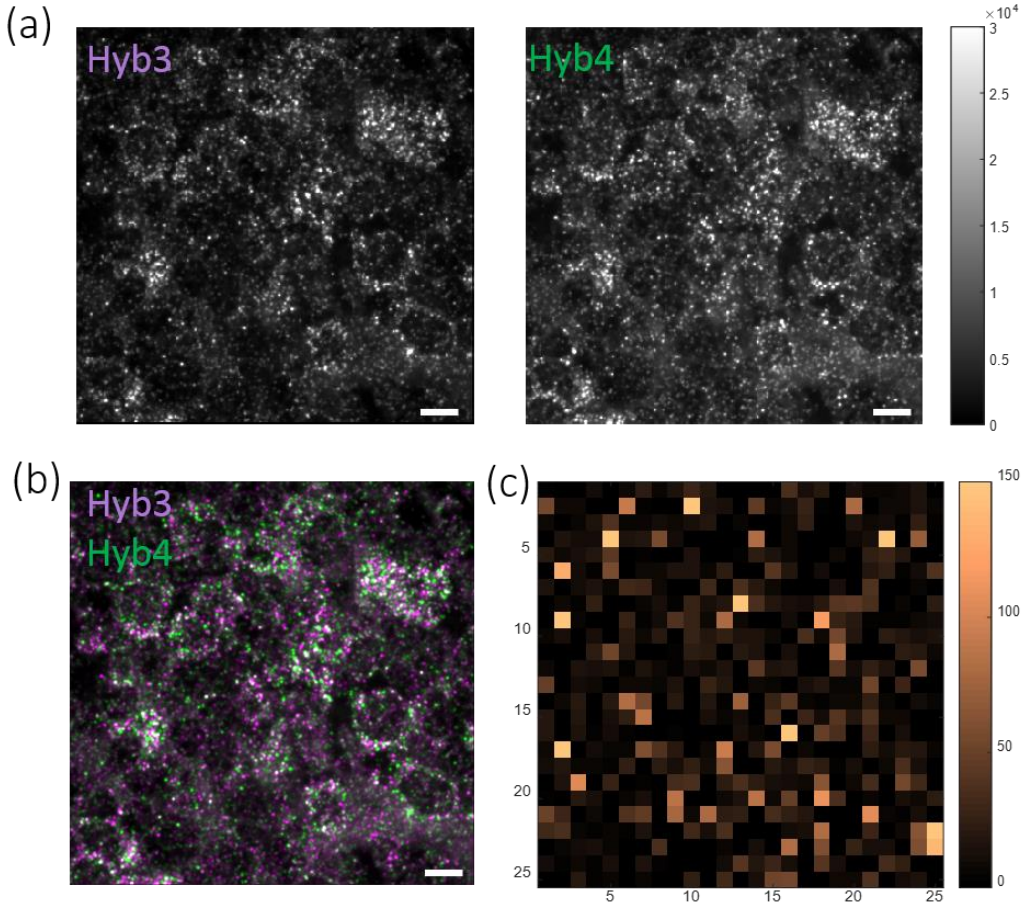


Supplementary Figure 11. Subcellular transcript quantification was performed on the subregions (*R*) from subsequent hybridizations. The *R* requires the two images to be registered. The size of *R* can be varied to create various *N* x *N* blocks in a cell. Each area is assigned to a color map for transcript abundance.



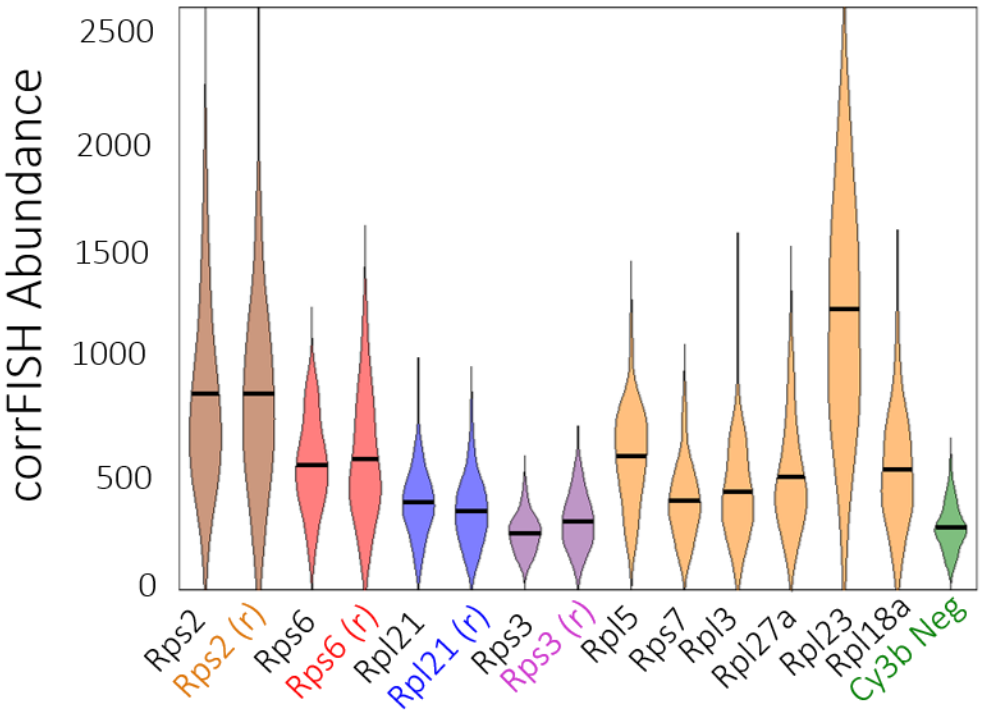
Supplementary Figure 12. Transcript quantification across various window sizes in simulations. (a) Hyb1 images with 1 μm^{-2} density of A and B molecules. (b) Hyb2 images with the same density of shared A and different C molecules. (c) Corresponding cross-correlation image in two-dimensions. (d) Cross-correlation in three-dimensional representation with a Gaussian fit (dashed). (e) Transcript detection results for 30 simulation runs in each case. Expected transcript counts (Ex), detected transcripts by corrFISH (Det), and the standard deviation (Std) of corrFISH results. These results suggest that corrFISH can be computed down to 10 x 10 pixels. However, accuracy of transcript detection is reliable after 16 x 16 pixels with significant variability. After 32x32 pixels, precision gets more accurate with lower standard deviations.

corrFISH quantification in tissues by spatial analysis



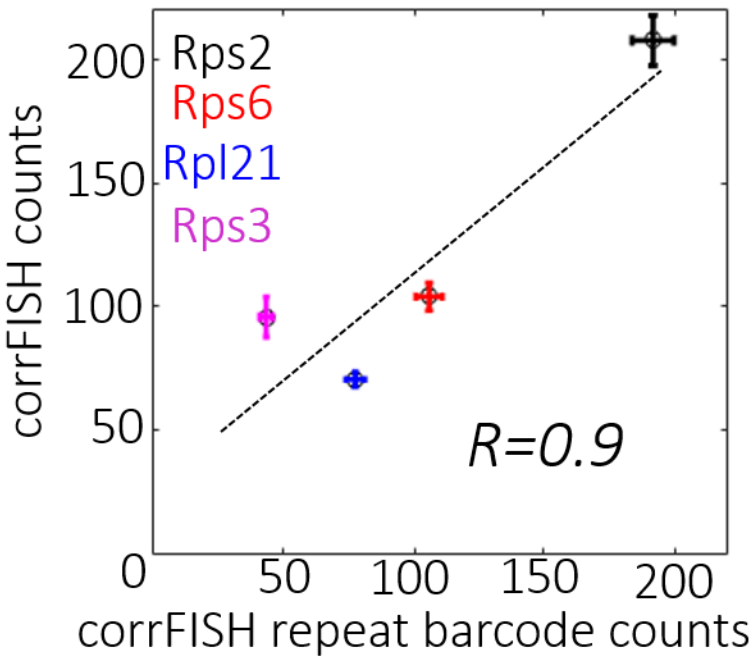
Supplementary Figure 13. Spatial analysis in tissue at single cell resolution. Subcellular analysis with 16 x 16 pixels window size (2.08 µm x 2.08 µm area) was used to measure *Rpl23* transcripts from Hybs 3 and 4. (a) Raw images for each hyb and (b) a merged image from both hybs. The resultant tissue image (400 x 400 pixels) that was processed by corrFISH across 25x25 ~ 625 sub-regions (c).

corrFISH in multiple layers of tissue section and transcript distributions



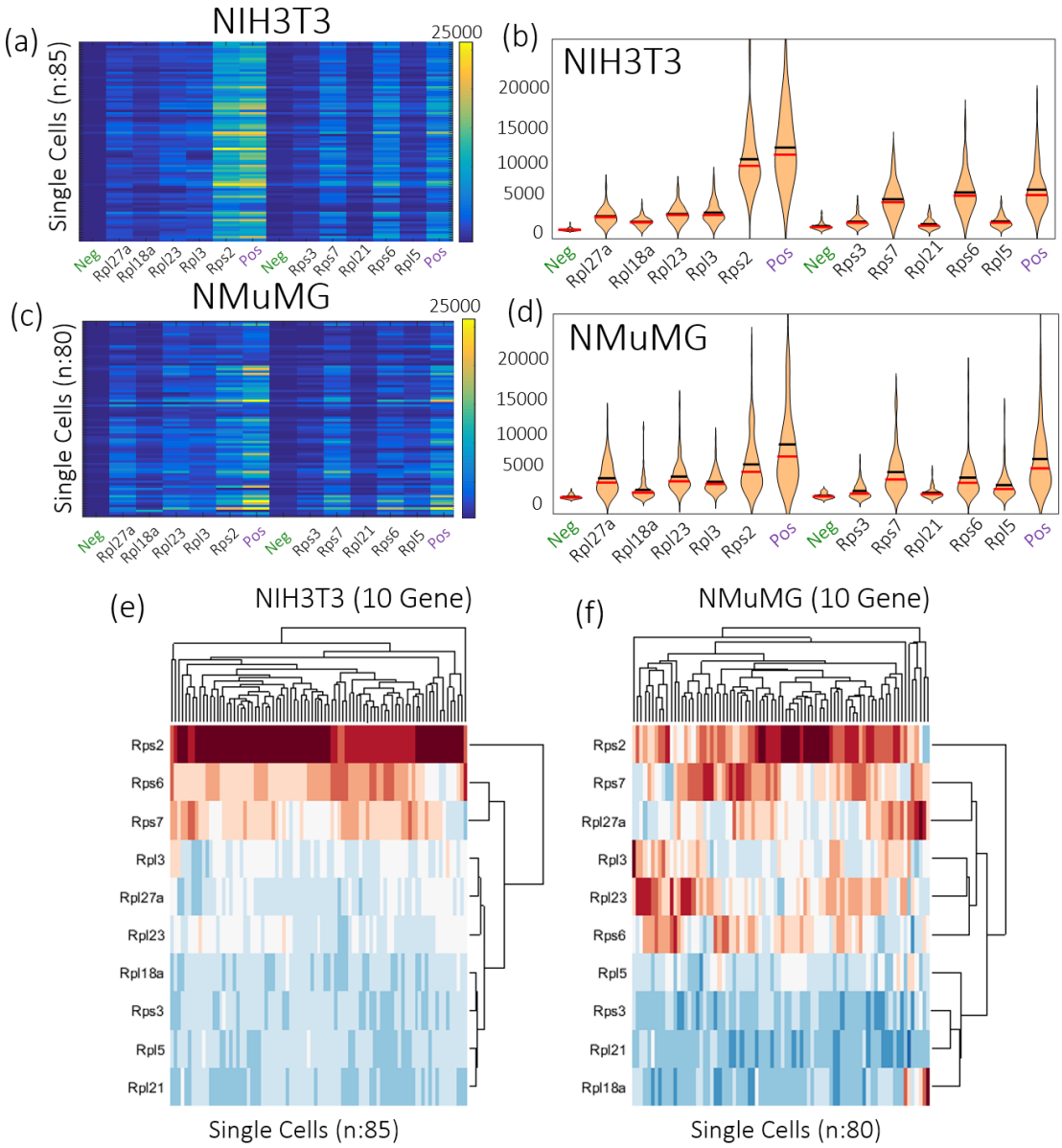
Supplementary Figure 14. Distributions of transcripts per cell measured by corrFISH experiments within a three µm thick optical layers of a thymus tissue section. Violin plots showing the mean (black) and distribution of transcript counts varying from 100 to 2,000 counts. Negative control was much lower than most of the genes, but it was higher compared to the cell culture control signal due to the high non-specific binding of FISH probes in tissue sections. Positive controls for corrFISH in tissue. Four ribosomal protein genes *Rps2* vs *Rps2 (r)*, *Rps6* vs *Rps6 (r)*, *Rpl21* vs *Rpl21(r)*, *Rps3* vs *Rps3 (r)* were detected by repeats. *n*=80 cells.

Repeat hyb analysis in corrFISH analysis in thymus



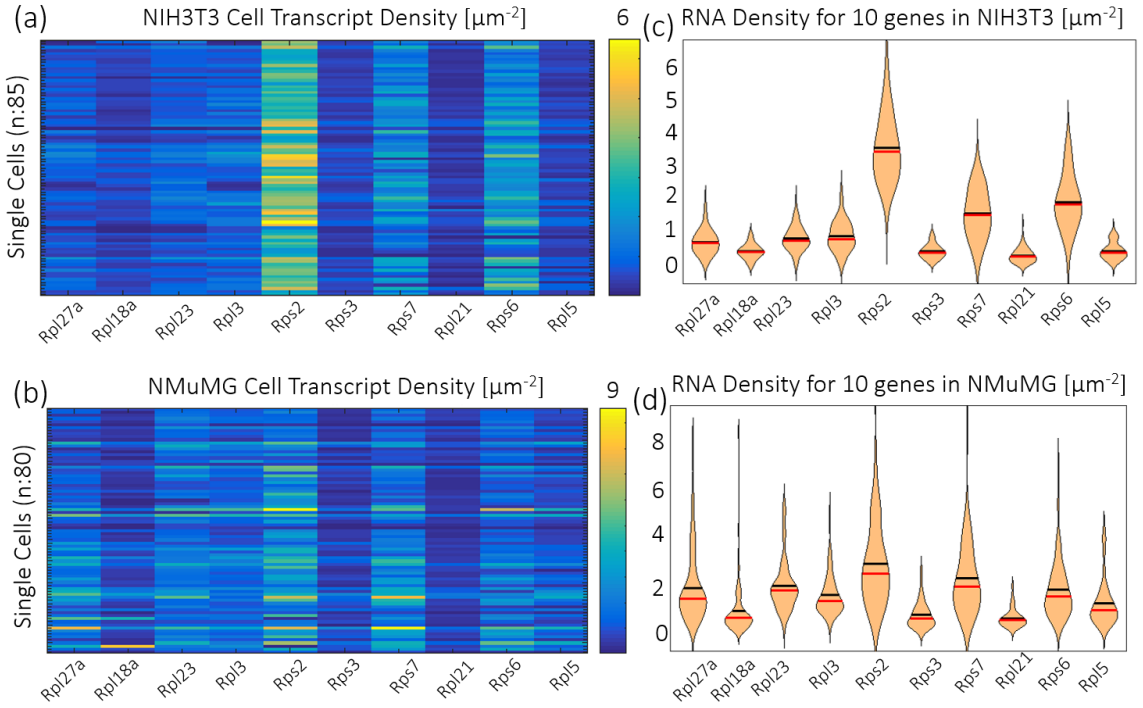
Supplementary Figure 15. corrFISH positive control within a 1 µm thick optical layers in a thymus tissue slice. *Rps2*, *Rps6*, *Rpl21*, and *Rps3* were measured twice as part of the barcoding scheme presented in Fig. 3b. The comparison of repeat measurements agreed well with each other, providing a linear fit with R=0.9 value and P<0.09 (Student's t-test). n=162 cells.

RNA copy number distributions and clustering in NMuMG and NIH3T3



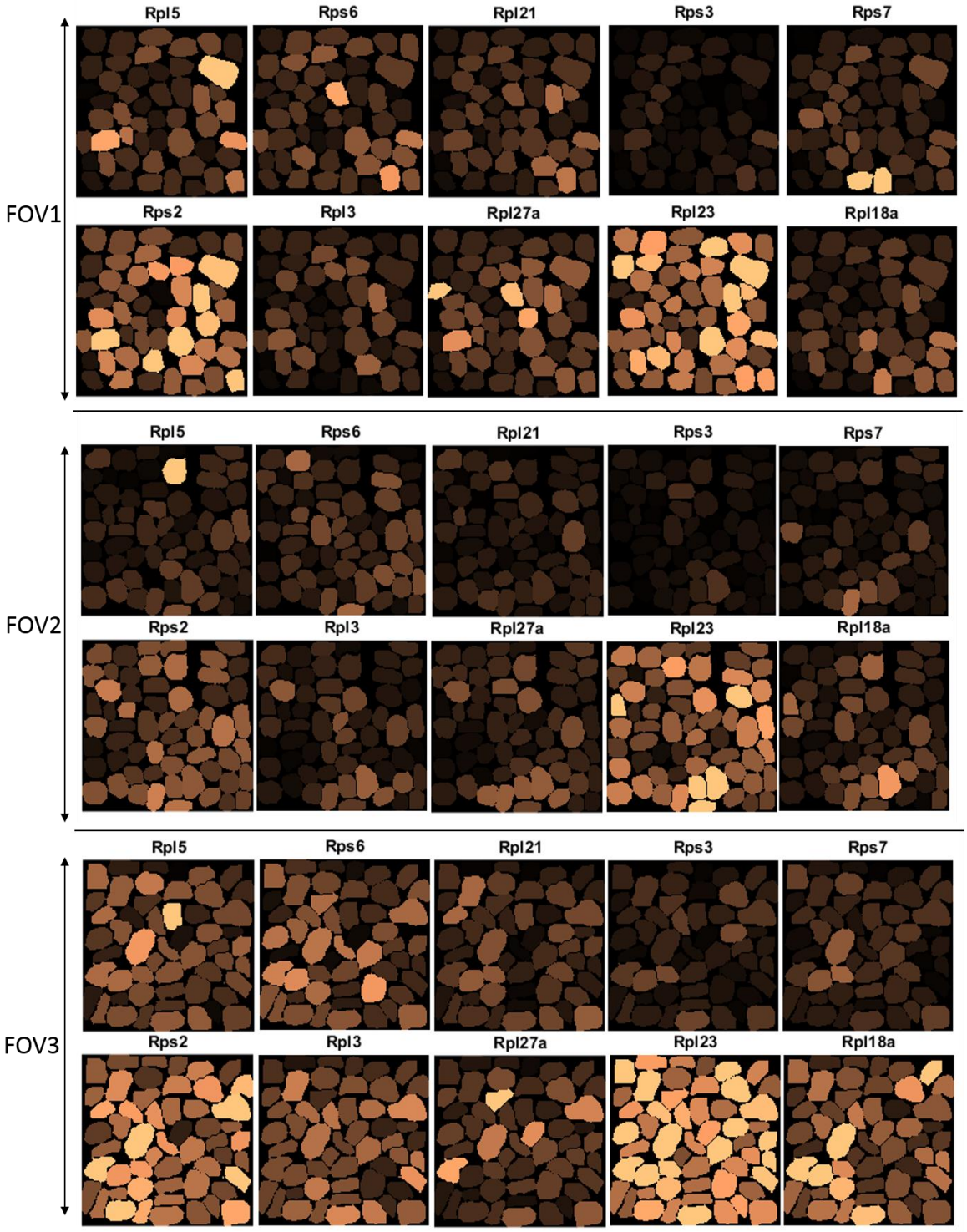
Supplementary Figure 16. Quantification of dense transcripts in distinct cell cultures: NIH3T3 and NMuMG. (a)-(c) Heat maps and (b)-(d) distribution plots for ribosomal protein transcripts in single cells determined by the image correlations, for *n*=85 cells and *n*=80 cells, respectively. Independently clustered gene expression of ten ribosomal proteins in (e) NIH3T3 and (f) NMuMG cells. These curves suggest combinatorial regulation of *Rps2*, a high copy transcript, with other transcripts such as *Rps7*, *Rpl27a*, *Rpl3* and *Rpl6*. The two each cell lines exhibit distinct co-expression patterns.

RNA density distributions in NMuMG and NIH3T3 cells



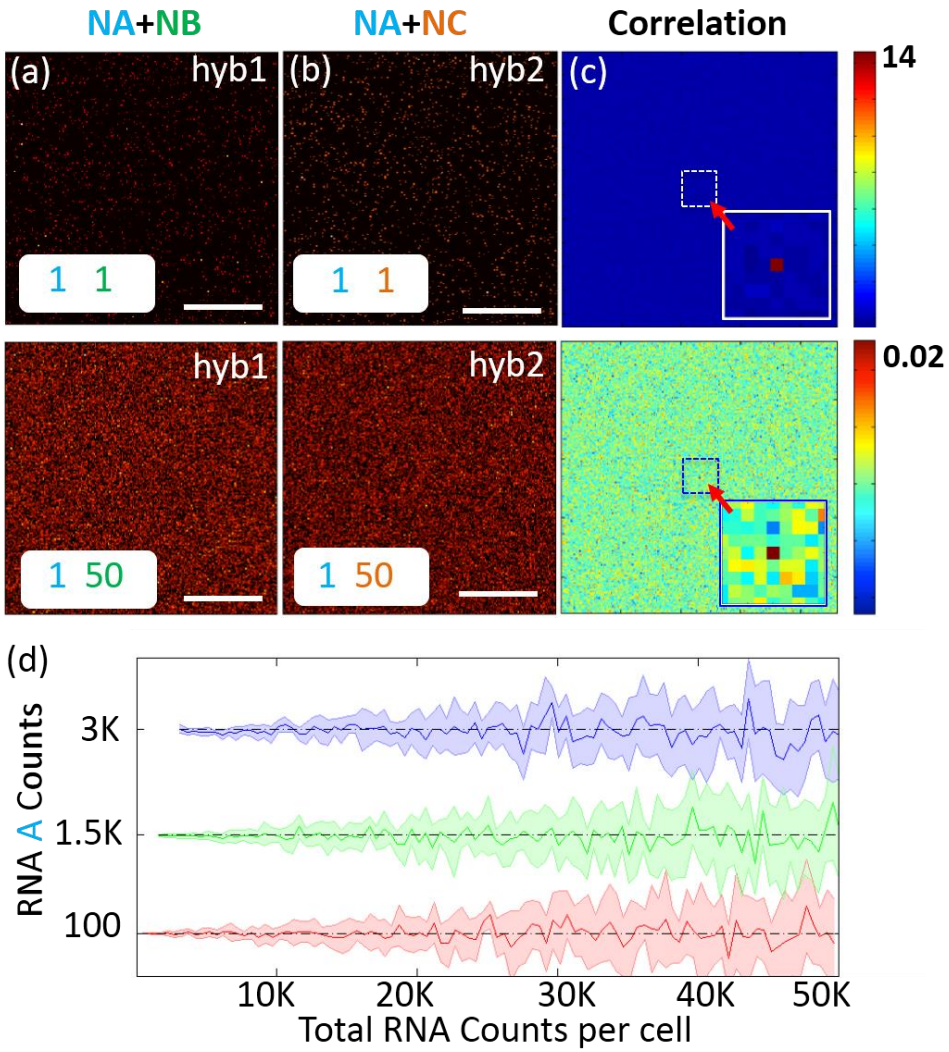
Supplementary Figure 17. Single cell transcript density heat map for ten ribosomal protein gene in (a) NIH3T3 and (b) NMuMG cells. Corresponding RNA density distributions in (c) and (d), respectively.

Spatial transcript distribution maps in three locations of a thymus tissue section.



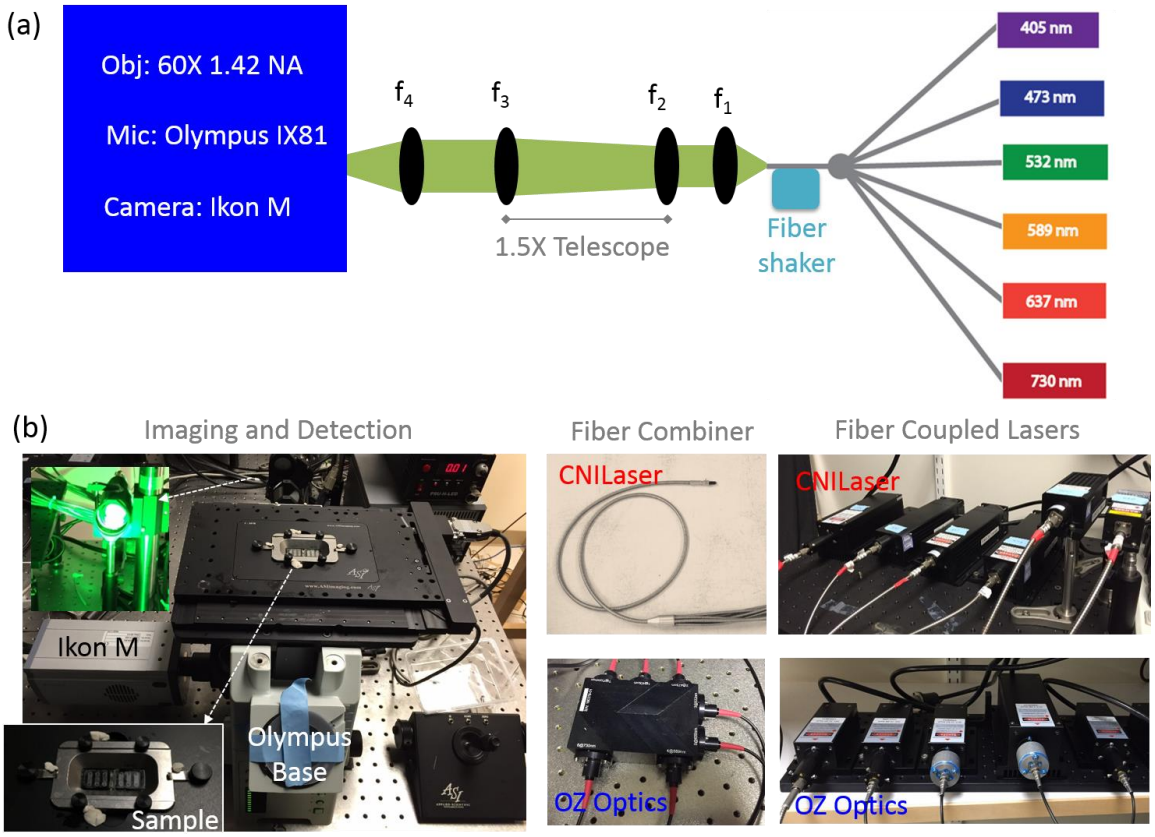
Supplementary Figure 18. Spatial transcript maps for ten ribosomal proteins in single cells within three different locations (FOV1, FOV2, and FOV3 are the positions presented in **Fig. 3c and 3d**) within a thymus section. All the images were contrasted the same to have maximum at 400 counts.

Image correlation analyses to improve the density of transcript detection in super resolution microscopy



Supplementary Figure 19. Correlation analysis applied to the images simulated to correspond to super resolution microscopy images. (a-b) Simulated FISH images with 50 nm PSF (to mimic super resolution data) were created for hybridizations 1 and 2. Transcript A was shared in between the two while transcripts B and C were uncorrelated. (c) Cross-correlation results of hyb 1 and hyb 2 images were computed for a 30 µm x 30 µm cell. Correlation peak appeared as a single delta function as PSF was smaller than the pixel size (130 nm). In comparison to the Supplementary Figure 1, detection of 1 μm^{-2} A molecule in the presence of other 50 μm^{-2} B or C molecules provided strong correlation peak that was much higher than the background. (d) RNA A counts of 100, 1,500, and 3,000 were quantified within a total RNA of up to 50,000. In comparison to the density presented in Figure 1g, 0.2% of the total transcripts were estimated within accurate mean values. Super resolution images improved the density more than 4-5 fold compared to the wide field microscopic images.

Experimental setup for a simplified single molecule-imaging microscope using a fiber combiner and lasers



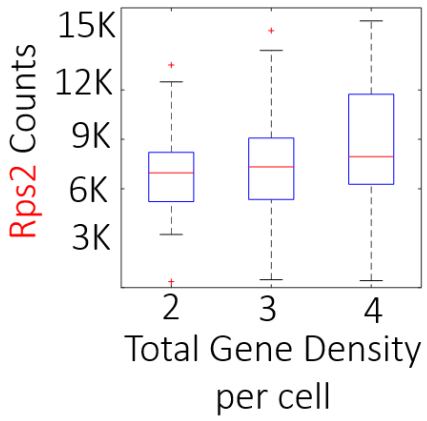
Supplementary Figure 20. A fiber illumination based wide field fluorescence microscope for single molecule imaging experiments. This approach alleviated the need for optical alignment in experimental set up for multi-color illumination of single cells.
(a) Schematic of experimental set up. A simple fiber combiner (1 to 7) was used to collect the light from each laser. The combined output was then collimated using a convex lens (f_1~200 mm). The collimated beam was then expanded 1.5 fold using a 2-lens telescopic system (f_2~200 mm and f_3~300 mm). The output fiber was then vibrated by a fiber shaker to avoid speckle issues, creating homogenous light output. Another lens was then used to focus the expanded beam at the back aperture of the 60X 1.42 NA objective lens. Olympus IX81 microscope base was used for imaging experiments. An additional 1.6X was used to collect single molecule signal, corresponding to 96X total magnification. An Andor Ikon-M camera was also used to enable single molecule sensitivity for FISH experiments. (b) Hardware components of the presented platform. Imaging and detection through Ikon M camera, Olympus IX81, convex lenses. Sample holder was shown for long term imaging experiments. Two kinds of fiber combiners were used to collect illumination. One is from CNILaser 7 to 1 tapered fiber that collected the light from fiber-coupled lasers through 400 µm multi mode laser. Another one is from OZ Optics 6 to 1 fiber multiplexer through one coupling box that also had inputs from fiber coupled lasers via 50 µm multi mode lasers.

Detection of Rps2 gene within total increasing transcript density

(a)

Gene	Color	H1	H2	H3	H4
Rps2	A647	1	1	1	1
Rps7	A647	0	1	1	1
Rpl3	A647	0	0	1	1
Rps3	A647	0	0	0	1

(b)

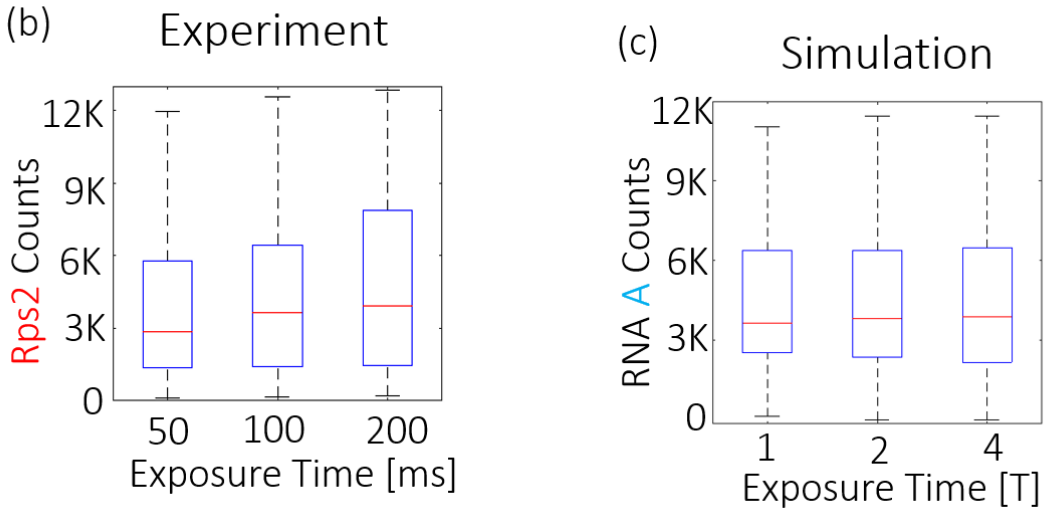


Supplementary Figure 21. Increasing total gene density do not affect accuracy in detecting *Rps2*. (a) Barcoding scheme for target 1 gene within 2, 3, and 4 total genes. (b) Corresponding boxplots for the *Rps2* counts per cell in NIH3T3 cells (*n*=49 cells). The counts are obtained by correlating H1 with H2, H3 and H4.

Detection of Rps2 gene at different exposure times

(a)

Gene	Color	H1 50 ms	H1 100 ms	H1 200 ms	H2 50 ms	H2 100 ms	H2 200 ms
Rps2	A647	1	1	1	1	1	1
Rps7	A647	1	1	1	0	0	0
Rpl27a	A647	1	1	1	0	0	0
Rpl3	A647	1	1	1	0	0	0
Rpl21	A647	1	1	1	0	0	0
Rps3	A647	0	0	0	1	1	1
Rpl18a	A647	0	0	0	1	1	1
Rpl23	A647	0	0	0	1	1	1
Rpsa	A647	0	0	0	1	1	1



Supplementary Figure 22. Detection of *Rps2* with different camera settings. Specifically, exposure time was adjusted to 50, 100, and 200 milliseconds**.** (a) Barcoding scheme for five genes to five genes, with only *Rps2* gene shared. (b) Bar plots for detection of *Rps2* with corresponding exposure times in NIH3T3 cultures (*n*=40 cells). (c) Similar conditions from simulated data on 40 cells.

Supplementary Note

Probe preparation. Probes were designed and prepared based on previous protocol[1].

Total gene density experiment in cultures. To evaluate the detection accuracy of corrFISH, we quantified one gene (*Rps2*) within the total of two, three, and four genes (**Supplementary Fig. 21**). The mean value of *Rps2* has not considerably changed (less than 5%) while the distribution got less than 2 fold wider. Once again, similar to the results in Fig. 1g, corrFISH provided accurate results with less than a factor of 2 of the mean value or the deviation.

Exposure time experiment in cultures. Next, we tested out the effect of various imaging conditions such as camera settings on the corrFISH quantification. In the experiments, we varied the exposure time per FISH image from 50 ms, 100 ms, and 200 ms (**Supplementary Fig. 22**). For these parameters, the mean value of *Rps2* target transcripts has not altered much (less than 5%) while the standard deviation slightly increases due to the appearance of new single molecules over the background at higher exposure times. Theoretically, signal levels are while computing the correlations. Thus, these experimental results confirmed analytical expectations.

Repeat hybridization experiments in cultures. One additional round of hybridization was performed for positive control. The positive control (hyb 5), where the probes used for hyb 3 is repeated, showed high correlation values with hyb 3 (**Supplementary Fig. 16**).

Single molecule FISH comparisons. We computed number of transcripts from hybridization images with one gene labeled at a time for comparisons to the corrFISH barcoding. For quantitation, we measured amount of connected dots in the Laplacian of Gaussian (LoG) filtered single molecule FISH images, yielding the total number of transcripts localized and separable within the image. Alternatively, the centroid of Gaussian fit function on to a single molecule dot can be used to measure the exact localization position of the object for enumeration of total RNA molecules. For single molecule integrated intensity analysis, we summed the intensity pixel values of a single molecule FISH image over the mask area of a cell.

Additional tissue analysis. We subtracted a blank image (without sample) from the FISH images to avoid illumination artifacts such as non-uniform interference patterns from the Yokogawa box.

Statistical analysis of plots. We used Excel and Origin to perform the "Student's *t*-test analysis" on the plots showing our correlation FISH results.

- For For Figure 1h of the manuscript, we obtained Pearson Correlation of 0.8314 with a functional form of y=0.7367x+1703.7. Based on n=43 observations (number of cells), analysis of variance (ANOVA) yields *F significance* ~ 5x10-12, suggesting accuracy of results and supporting correlations of two separate corrFISH measurements. A *t*-test analysis provides *t statistics* of 3.34, corresponding to *P*-value of 0.0017. These *F* and *P* values showed that our Rps2 gene measurements by "M1: H1*H2" and "M2: H2*Rps2 from H3" agreed well with each other.

- For Figure 1i, smFISH counting efficiency is variable based on the transcript densities. While the entire dynamic range may not exhibit a uniform correlation, the lower end of the curve

exhibits relatively linear relation of smFISH to corrFISH counting. Thus, we focused onto the range of 0-3,000 transcripts for our statistical analysis. The regression analysis provided Pearson Correlation of 0.8638 with a functional form of y=0.8458x+693.54. Using 11 observations (number of cells) in this range for a student's t-test and ANOVA analysis, we obtained an *F significance* < 0.0006, *t statistics* ~ 5.1440, and *P* value of 0.0006. *P and F insignificance* validated the importance of our results. Two data points were omitted in this analysis due to their high cook's distance.

- For Figure 2d, smFISH and corrFISH comparisons exhibit an exponential trend with anR-Square value of 0.91. Using functional form y=a*(1-exp(-b*x), we obtained y=2813.75*(1-exp(-0.0004071*x), where x is corrFISH value and y is the corresponding smFISH value for a data point. ANOVA analysis provided insignificant *F* probability <0.0009, providing a successfully exponential curve fitting in our comparisons.

- For Supplementary Figure 15, we obtained a Pearson correlation of 0.90 with a functional form of y=0.8645x+29.09. Using n=162 observations (number of cells), ANOVA analysis provided an F probability <0.094 which is a small value to support correlations of our repeat barcodes. Similarly, a t-test statistical analysis yielded t statistics of 3.0448, corresponding to P-value of 0.0930, which is smaller than our typical test parameter α=0.10. These negligible values of P-value and F-insignificance verified the detection of transcripts within a tissue section.

Theory

Spatial Image Correlations. Abundance of a molecule in single cells is quantified from image correlations. Two approaches are used to perform correlations, one in spatial domain and another in Fourier domain. The spatial domain is implemented by shifting one of the images and multiplying the pixel values to create a correlation matrix (Equation 1).

$$G_{12}(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} H_1\,(m,n) \times \overline{H_2}(m-i, n-j) \tag{1}$$

where M and N are dimensions of images, H_1 and H_2 are the images from hybridizations, G is the correlation matrix, i and j are the spatial lag variables.

Spatial domain method works accurately for image correlation, however, it requires extensive multiplications making it slow processing scheme. Thus, a Fourier domain approach is used to take the Fourier Transform of both images, multiply both images once, and then take inverse transform back to the spatial domain (Equation 2), providing much faster results.

$$G_{12} = \mathcal{F}^{-1}\{\mathcal{F}(H_1)\mathcal{F}^*(H_2)\} \tag{2}$$

where $\mathcal{F}$ is the Fourier transform operation, and H_1 and H_2 are the images, and G_{12} is the correlation matrix.

To correct for the deviations in the pixel brightness of two images, the image correlation function is normalized and an offset value is subtracted to compute the correlation function[2] (Equation 3).

$$G_{12} = \frac{\mathcal{F}^{-1}\{\mathcal{F}(H_1)\mathcal{F}^*(H_2)\}}{\langle H_1\rangle\langle H_2\rangle} - 1 \tag{3}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform operation, F denotes Fourier transform operation, H_1 and H_2 are the images, and G is the correlation matrix. $\langle H_1\rangle$ and $\langle H_2\rangle$ are the mean value of entire H_1 and H_2 images, respectively.

Derivation of correlation amplitudes. Autocorrelation of each channel is used to measure to total number of molecules in a given color channel. In Equations 1 to 3, A and B images are taken as the same, as the autocorrelation implies correlation of the image by itself. As in Fig 2, there are N_A molecules of A and N_B molecules of B in hybridization 1 image, then total signal for the image in $H_1(x,y)=C_1\ (N_A(x,y)+N_B(x,y))$ where C_1 is the intensity coefficient for that imaging channel, which depends on light intensity, exposure time, dye brightness etc. Similarly, for the second image containing both A and C molecules, then that image $H_2(x,y)=C_2\ (N_A(x,y)+N_c(x,y))$ where C_2 is the intensity coefficient for the 2nd imaging channel.

$$G_{12} = \frac{\langle (H_1 - \langle H_1\rangle)(H_2 - \langle H_2\rangle)\rangle}{\langle H_1\rangle\langle H_2\rangle} = \frac{\langle H_1 H_2\rangle}{\langle H_1\rangle\langle H_2\rangle} - 1 = \frac{C_1C_2\langle (N_A+N_B)(N_A+N_C)\rangle}{C_1C_2(N_A+N_B)(N_A+N_C)} - 1$$

$$= \frac{\langle N_A N_A\rangle + \langle N_A\rangle\langle N_B\rangle + \langle N_A\rangle\langle N_C\rangle + \langle N_C\rangle\langle N_B\rangle}{(N_A+N_B)(N_A+N_C)} - 1 \tag{4a}$$

Bracket denotes integration over the x,y coordinates. Since only N_A (x, y) are correlated between the two images, only $\langle N_A N_A\rangle$ correlation remains within the integral, whereas $\langle N_A N_B\rangle = \langle N_A\rangle\langle N_B\rangle$. If $N_A(x, y)$ is Poisson distributed in space, then $\langle N_A N_A\rangle = \langle N_A\rangle + \langle N_A\rangle^2$(Here the brackets are expectation values). Eqn 4 then becomes

$$G_{12} = \frac{\langle N_A\rangle}{(N_A+N_B)(N_A+N_C)} \tag{4b}$$

Similarly,

$$G_{11} = \frac{1}{(N_A+N_B)}$$

$$G_{22} = \frac{1}{(N_A+N_C)} \tag{5}$$

where G_{11} and G_{22} are the amplitudes of autocorrelation functions from hybs 1 and 2, respectively. The cross-correlation across color channels enables detection of c, shared molecules across these images by the multiplication of autocorrelation amplitudes, which is the transcript abundance per point spread function (PSF) area (Equation 6)

$$\langle d_{12} \rangle = \frac{G_{12}}{G_{11} G_{22}} \tag{6}$$

where d_{12} is the transcript copy number barcoded by Hyb 1 and 2 per PSF area to be detected by corrFISH. G_{12} is the amplitude of the cross-correlation function and will be divided by G_{11} and G_{22} autocorrelation amplitudes of hyb1 and hyb2 images.

$$\langle N_{12} \rangle = \frac{\langle d_{12} \rangle}{A_{PSF}} x A_{Cell} \tag{7}$$

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 abundance of transcripts per PSF area from Equation 2. A_{PSF} is the PSF area ($\pi\ x\ PSF_{width}{}^2$), where PSF width is 0.3 μm. A_{Cell} is the area of the single cell image ($N\ x\ p^2$), 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.

Higher order correlations for transcript quantification. A common molecule type across many images is also computed using a similar strategy[3-5]. For 3 hybridizations, the cross correlation of three images are computed and then divided by autocorrelation amplitudes of each images (Equations 8 and 9).

$$\langle d_{123} \rangle = \frac{G_{123}}{G_{11} G_{22} G_{33}} \tag{8}$$

$$\langle N_{123} \rangle = \frac{\langle d_{123} \rangle}{A_{PSF}} x A_{Cell} \tag{9}$$

Higher order correlations of N images follow the same general scheme of molecule quantification (Equations 10 and 11). The amplitude of cross correlations of N images is divided by individual autocorrelation amplitudes.

$$\langle d_{12\ldots N} \rangle = \frac{G_{12\ldots N}}{G_{11} G_{22} \ldots . G_{NN}} \tag{10}$$

$$\langle N_{12\ldots N} \rangle = \frac{\langle d_{12\ldots N} \rangle}{A_{PSF}} x A_{Cell} \tag{11}$$

References

1. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. *Nat. Methods* **11**, 360–361 (2014).
2. Costantino, S., Comeau, J. W. D., Kolin, D. L. & Wiseman, P. W. *Biophys. J.* **89,** 1251–1260 (2005).
3. Kettling, U., Koltermann, A., Schwille, P. & Eigen, M. *Proc. Natl. Acad. Sci.* **95,** 1416–1420 (1998).
4. Schwille, P. *Cell Biochem. Biophys.* **34,** 383–408 (2001).
5. Heinze, K. G., Jahnz, M. & Schwille, P. *Biophys. J.* **86,** 506–516 (2004)