

Supplementary Information for

Large-scale image-based transcriptomics in thousands of single human cells at single-molecule resolution

Nico Battich^{1,2,3}, Thomas Stoeger^{1,2,3}, Lucas Pelkmans¹

¹Faculty of Sciences, Institute of Molecular Life Sciences, University of Zurich,
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.

²Life Science Zurich Graduate School, Ph.D. program in Systems Biology.

³ contributed equally

Correspondence should be addressed to L.P. (lucas.pelkmans@imls.uzh.ch)

Contents

Supplementary Protocol: High-throughput image-based transcriptomics using bDNA sm-FISH.	3
Supplementary Note 1: Calculation of signal-to-noise ratios.	8
Supplementary Note 2: Experimental setup and calculation of percentage of spots detected by two probe set types and technical noise contribution.	9
Supplementary Note 3: Robust spot detection for image-based transcriptomics.	10
Supplementary Note 4: Calculation of fraction of cells with spot counts above background and mean spot per cell correction.	13
Supplementary Note 5: Identification of cellular patterns of mRNA localization.	14
Supplementary Note 6: Preparation of multivariate transcript readouts for clustering and feature selection.	17

Supplementary Tables.	21
References.	30
Supplementary Figures.	31

Supplementary Protocol: High-throughput image-based transcriptomics using bDNA sm-FISH

Cells were seeded as described in Online Methods in complete medium supplemented with Penicillin/Streptomycin (Gibco). Cells were fixed in 4% paraformaldehyde for 30 min at RT. Incubation of primary probe pairs was done for 3 hrs, while for pre-amplifiers, amplifiers and label probes for 1 hr each. All hybridization steps were done at 40°C in a Liconics rotating incubator to avoid plate positional effects. After hybridization reactions, cells were stained for 10 min with 0.2µg/ml DAPI in PBS and then incubated for 5 min in 1ng/µl of Alexa Fluor® 647 carboxylic acid, succinimidyl ester (Invitrogen) in carbonate buffer (1.95ml of 0.5M NaHCO₃, 50µl of 0.5M Na₂CO₃ in and 8ml of water for 10ml of buffer) for detection of the cell outline. For image-based transcriptomics, prior to fixation cells were incubated in 0.5µM MitoTracker® Red CMXRos for 45 min at 37°C for detection of mitochondria. All steps in the protocol were automated using the EL406 washer-dispenser from BioTek and a Bravo liquid handling platform from Agilent Technologies.

High-throughput in situ RNA hybridization: Main protocol

Experiment Title:

Date:

Notes

- 1 This protocol assumes processing of three 384-well plates in sequential order.
- 2 Every aspiration step was done in the EL406 BioTek washer-dispenser so that 15ul residual volume were left in the well.
- 3 Tips used for dispensing and mixing with Bravo platform were 70ul Tips (# 19133-212). Tips were freshly discharged with a Milty Zerostat3 not more than 1h before immediate use.
- 4 This protocol is based upon Affymetrix's protocol: QuantiGene ViewRNA HC Screening User Manual, RevD

- 5 Few remaining manual steps could also be fully automated.
PreHyb plates can be refilled during the assay. This refilling is not
- 6 indicated.
Prepare two plates of PreHyb to reduce effect of unexpected problems.
Prepare protocols for machines before starting the assay (and double check processed
- 7 volumes manually).
- 8 For fixation, fix all wells of the plate (not only the actually used wells)

Abbreviations

RT	Room Temperature
PBS	Phosphate buffer saline
PFA	Paraformaldehyde
SuccEst	Alexa Fluor® 647 carboxylic acid, succinimidyl ester (Invitrogen)

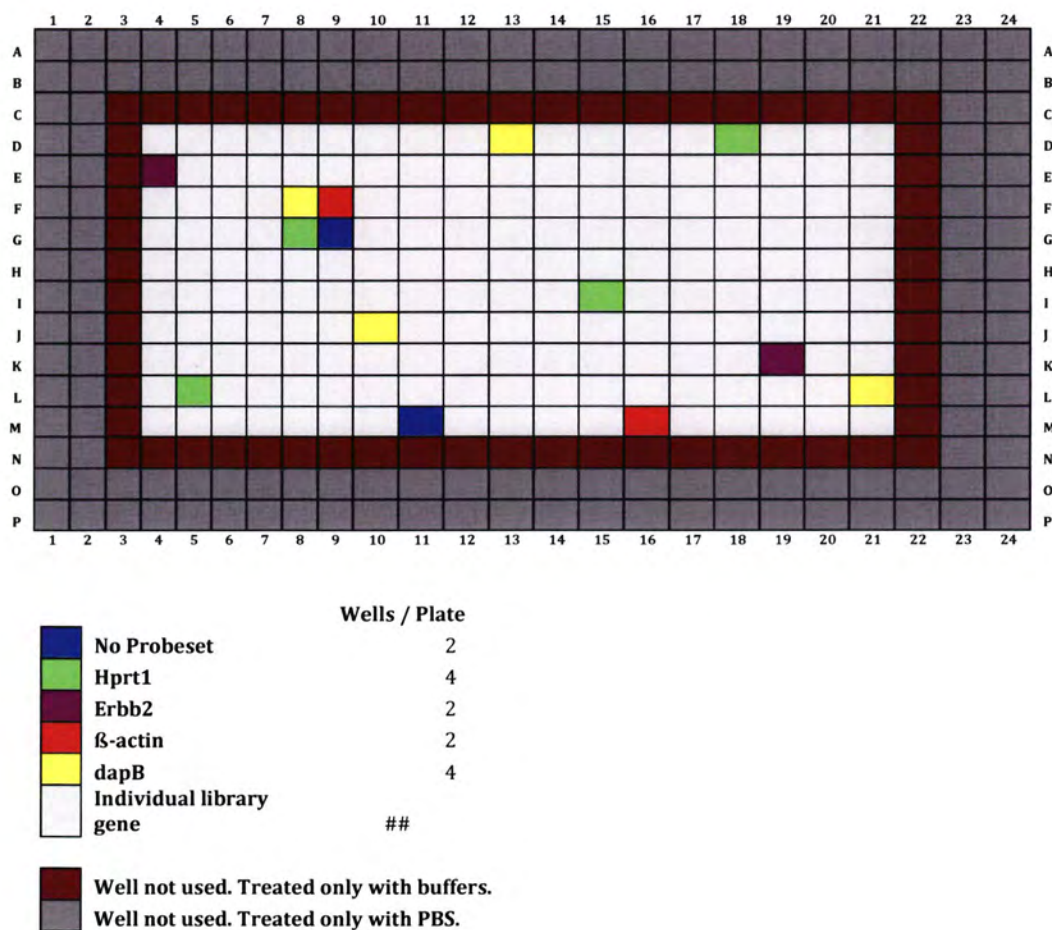
Step	Hour	Min	Plate	Name	Description	Check	Comments
1			N/A	Reagents	Warm Reagents: PreHyb, PS_Diluent, Amp_Diluent, LP_Diluent, PBS (for PreHyb plates only) for 30min 40C		
2			N/A	Reagents	Prepare 8% PFA ad leave at RT.		
3			N/A	Reagents	Prepare Wash Buffer ad leave at RT.		
4			N/A	Reagents	Prepare Probe Sets leave at RT.		
5			N/A	Reagents	Make sure PreHyb is at 40C. Prepare 2 plates and store at 40C.		
6			N/A	Reagents	Prepare Pre Amp leave at RT.		
7			N/A	Reagents	Prepare Amp leave at RT.		
8			N/A	Reagents	Prepare Label Probe leave at RT.		
9			N/A	Equipment	Start Biotek EL406 and test functionality, 1ul Cassette		
10			N/A	Equipment	Start Bravo (and initialize Liconic)		
11			N/A	Break	Have Breakfast		
12			N/A	Reagents	Prepare Mitotracker in Serum Free Medium. Use within 1min.		OPTIONAL
13			All	Mitotracker	Aspirate medium, dispense Mitotracker (15ul), Incubate@37C for 45 min.		OPTIONAL
14			All	Cell Fixation	Wash 2x with PBS (80ul), fix plate with 8% PFA (15ul) for 30 min at RT.		
15			All	Cell Fixation	Aspirate PFA and wash 3x with PBS.		
16			N/A	Equipment	Remove First and Last tubes of Biotek EL406 1ul cassette.		
17			1	Cell Perm.	Aspirate PBS, dispense Detergent Solution (15ul) and incubate for 3 min at RT.		
18			1	Cell Perm.	Wash 2x with PBS.		
19			1	Protease	Aspirate PBS, dispense Working Protease (15ul), incubate 10 min at RT with lid closed.		
20			1	Protease	Wash 5x with PBS.		
21			1	Protease	Aspirate PBS, dispense Protease Stop Buffer (15ul).		
22			1	Protease	Mix by pipetting up and down twice		
23			1	Protease	Aspirate Protease Stop Buffer , dispense Protease Stop Buffer (15ul).		
24	0	0	1	Probe Set	Dis pence Working Probe Set with Bravo (15ul). Place Plate in incubator		
25	0	10	2	Cell Perm.	Aspirate PBS, dispense Detergent Solution (15ul) and incubate for 3 min at RT.		

26			2	Cell Perm.	Wash 2x with PBS.		
27			2	Protease	Aspirate PBS, dispense Working Protease (15ul), incubate 10 min at RT with lid closed.		
28			2	Protease	Wash 5x with PBS.		
29			2	Protease	Aspirate PBS, dispense Protease Stop Buffer (15ul).		
30			2	Protease	Mix by pipetting up and down twice		
31			2	Protease	Aspirate Protease Stop Buffer , dispense Protease Stop Buffer (15ul).		
32			2	Probe Set	Dispense Working Probe Set with Bravo (15ul). Place Plate in incubator		
33	0	50	3	Cell Perm.	Aspirate PBS, dispense Detergent Solution (15ul) and incubate for 3 min at RT.		
34			3	Cell Perm.	Wash 2x with PBS.		
35			3	Protease	Aspirate PBS, dispense Working Protease (15ul), incubate 10 min at RT with lid closed.		
36			3	Protease	Wash 5x with PBS.		
37			3	Protease	Aspirate PBS, dispense Protease Stop Buffer (15ul).		
38			3	Protease	Mix by pipetting up and down twice		
39			3	Protease	Aspirate Protease Stop Buffer , dispense Protease Stop Buffer (15ul).		
40			3	Probe Set	Dispense Working Probe Set with Bravo (15ul). Place Plate in incubator		
41	1	20	N/A	Equipment	Change Cassette of BioTek to 5ul Cassette		
42	1	25		Break	Lunch		
43	2	50	1	Probe Set	Aspirate Working Probe Set , do 3x Wash Buffer at RT followed by 30 sec incubation.		
44			1	Pre Amp	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
45			1	Pre Amp	Dispense Working Pre Amp (15ul) with Bravo by quadrant. Place plate in the incubator.		
46	3	40	2	Probe Set	Aspirate Working Probe Set , do 3x Wash Buffer at RT followed by 30 sec incubation.		
47			2	Pre Amp	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
48			2	Pre Amp	Dispense Working Pre Amp (15ul) with Bravo by quadrant. Place plate in the incubator.		
49	4	0	1	Pre Amp	Aspirate Working Pre Amp , do 3x Wash Buffer at 40C followed by 30 sec incubation.		
50			1	Amp	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
51			1	Amp	Dispense Working Amp (15ul) with Bravo by quadrant.		
52	4	30	3	Probe Set	Aspirate Working Probe Set , do 3x Wash Buffer at RT followed by 30 sec incubation.		

53			3	Pre Amp	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
54			3	Pre Amp	Dispense Working Pre Amp (15ul) with Bravo by quadrant. Place plate in the incubator.		
55	4	50	2	Pre Amp	Aspirate Working Pre Amp , do 3x Wash Buffer at 40C followed by 30 sec incubation.		
56			2	Amp	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
57			2	Amp	Dispense Working Amp (15ul) with Bravo by quadrant.		
58	5	10	1	Amp	Aspirate Working Amp , do 3x Wash Buffer at RT followed by 30 sec incubation.		
59			1	LP	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
60			1	LP	Dispense Working Label Probe (15ul) with Bravo by quadrant.		
61	5	40	3	Pre Amp	Aspirate Working Pre Amp , do 3x Wash Buffer at 40C followed by 30 sec incubation.		
62			3	Amp	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
63			3	Amp	Dispense Working Amp (15ul) with Bravo by quadrant.		
64	6	0	2	Amp	Aspirate Working Amp , do 3x Wash Buffer at RT followed by 30 sec incubation.		
65			2	LP	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
66			2	LP	Dispense Working Label Probe (15ul) with Bravo by quadrant.		
67	6	20	1	LP	Aspirate Working Label Probe , do 3x Wash Buffer at RT followed by 30 sec incubation.		
68			1	LP	Wash 3x with PBS.		
69	6	50	3	Amp	Aspirate Working Amp , do 3x Wash Buffer at RT followed by 30 sec incubation.		
70			3	LP	Aspirate Wash Buffer, dispense PreHyb Buffer (15ul) with Bravo by quadrant, pipette up and down twice, aspirate PreHyb		
71			3	LP	Dispense Working Label Probe (15ul) with Bravo by quadrant.		
72	7	10	2	LP	Aspirate Working Label Probe , do 3x Wash Buffer at RT followed by 30 sec incubation.		
73			2	LP	Wash 3x with PBS.		
74			N/A	Reagents	Prepare DAPI.		
75	7	50	3	LP	Aspirate Working Label Probe , do 3x Wash Buffer at RT followed by 30 sec incubation.		

76			3	LP	Wash 3x with PBS.		
77	8	0	All	DAPI	Aspirate PBS, dispense Working DAPI reagent (70ul), incubate at RT for 10 min		
78			All	DAPI	Wash 3x with PBS.		
79	8	20	N/A	Reagents	Prepare SuccEst.		
80	8	30	All	SuccEst	Aspirate PBS, dispense Working SuccEst reagent (70ul), incubate at RT for 5 min		
81			All	SuccEst	Wash 3x with PBS.		
82	8	40	All	Cover	Cover all plates with sticky metal foil		
83	8	50	1	Microscope	Image plates (do not make new settings if tired)		

High-throughput in situ RNA hybridization: Plate layout example



Supplementary Note 1: Calculation of signal-to-noise ratios.

To obtain an accurate estimate of the signal per each individual spot including those in crowded areas we performed sub-pixel 2D Gaussian fitting with **Supplementary Equation 1**

$$f(x, y) = ae^{-\left(\frac{(x-x_0)^2}{2\sigma^2} + \frac{(y-y_0)^2}{2\sigma^2}\right)}$$

where x and y represent coordinates in the image, x_0 and y_0 represent the center of the spot, σ is the standard deviation of the Gaussian distribution which is assumed to be the same in both dimensions, a is the amplitude of the Gaussian and e is the Euler's number. First, 100 detected spots were randomly picked per condition and a neighborhood of 50 pixels around the spot cropped for optimization. Each pixel P of every spot was then subdivided into a 2D Euclidian grid of sub-pixels p_i such that a length of each sub-pixel was given by $\|p_i\| = \frac{\|P\|}{k}$, where $\|\cdot\|$ is a 2-norm, $k = 5$ gives a total number of 25 sub-pixels p_i per each pixel P in the cropped spot. The local background B_{loc} was subtracted and all sub-pixel intensity values normalized by dividing with the maximum intensity, or amplitude A , in the given spot so that a in s.eq.1 was set to one. Parameters for s.eq.1 fitting each spot were learned using the minimization method (COBYLA)¹ implemented as a part of the NLOpt non-linear optimization package (<http://ab-initio.mit.edu/nlopt>) by giving a constraint (**Supplementary Equation 2**)

$$S = \sum_{i=1}^N |r_i - m_i|, \quad S \rightarrow \min$$

where S is the optimization score, r_i and m_i are the real and modeled normalized intensity values for sub-pixel p_i and N is the total number of sub-pixels in the cropped

image.

The intensity brightness b_i of each sub-pixel p_i in each spot was then estimated by $b_i = m_i \cdot A$. The spot brightness after local background subtraction I_{lb} was then given by the integrating b_i for each spot, i.e. $I_{lb} = \sum_{i=1}^N b_i$. In order to estimate the signal-to-noise ratio SNr , we first computed the intensity brightness above the image background B_{im} , namely the mean intensity of pixels outside cells, for each sub-pixel p_i given by, $c_i = b_i + B_{loc} - B_{im}$. The signal-to-noise ratio snr_i of every sub-pixel b_i was then given by $snr_i = c_i / (B_{loc} - B_{im})$.

Supplementary Note 2: Experimental setup and calculation of percentage of spots detected by two probe set types and technical noise contribution.

Probe sets of type 1 and type 6 against *KIF11* and *ERBB2* transcripts were designed so that each probe pair of each type alternated along the transcript length (**Supplementary Fig. 3a**) to avoid differences in spot detection arising from possible hybridization biases between the 3' and 5' ends of the transcripts. bDNA FISH was carried out as described in the **Supplementary Protocol** and cells imaged in our high-magnification or the high-throughput set up. Type 1 probe pairs were labeled with AlexaFluor 546 tagged label probes and type 6 probe pairs with AlexaFluor 488 tagged label probes. Spots were detected for both probe types and the percentage of type 1 spots that were also detected with the type 6 probes set calculated. For the high-throughput setup we also performed cell segmentation and spot overlap was calculated for 5,289 and 5,391 single cells for probe sets targeting *KIF11* and *ERBB2*, respectively.

The contributions of technical noise and biological variability to the total single cell data variation was calculated as described by Elowitz *et al.*² for intrinsic, extrinsic and total noise, respectively.

Supplementary Note 3: Robust spot detection for image-based transcriptomics

Illumination correction of images

We exploited the large amount of images acquired per plate and per channel in this study to learn the illumination and signal gain differences in our field of view and specific imaging set-up. For each channel we calculated the mean intensity μ_i and the standard deviation σ_i of each pixel p_i in the field of view. We then derived an overall mean intensity $\bar{\mu}$ as well as the mean standard deviation $\bar{\sigma}$ of all pixels. To correct the illumination bias we performed per-pixel z-scoring as in **Supplementary Equation 3**

$$z_i = \frac{In_i - \mu_i}{\sigma_i}$$

where z_i is the z-scored value for pixel p_i and In_i is the original intensity value for pixel p_i in a given image. The corrected intensity value C_i for pixel p_i in an image was then calculated as shown in **Supplementary Equation 4**

$$C_i = z_i \cdot \bar{\sigma} + \bar{\mu}$$

Rescaling of images for spot detection

Image intensities were rescaled such that the image minimum, defined as the 0.01 quantile of the intensities, would become zero and the image maximum, defined as the

0.995 quantile of the intensities, would become one and all other values would scale accordingly allowing for negative values and values larger than one. To make the rescaling robust against very dim images without signal or images with unusually high intensities, plate-wide limits for the rescaling were derived from bDNA sm-FISH negative and positive controls included on each plate. A total of 98 images of each *dapB* and the *HPRT1* controls were randomly selected to serve as a reference. The lower limit of the minimal intensity was set to the 0.1 quantile of the minimal intensities of the *dapB* control. The upper limit of the minimal intensity was set to the 0.8 quantile of the maximal intensities of the *dapB* control. The lower limit of the maximal intensity was set to the 0.4 quantile of the maximal intensities of the *HPRT1* control. The upper limit of the maximal intensity was set to the 0.8 quantile of the maximal intensities of the *HPRT1* control (arrows in **Supplementary Fig. 7b**). Rescaling parameters that fell outside of the boundaries were set to the respective lower or upper limits.

Spot detection

For spot detection we performed a Laplacian-of-Gaussian (LoG) filter to the rescaled images and defined all objects above a certain threshold as spots. Note that the LoG has been successfully applied by for enhancing mRNA spots before³. The filter size was of five pixels, which corresponds to 5x162.5nm (812.5nm). We used one global threshold for all images of one plate after ensuring that corrected intensities of individual images were comparable. The specific threshold value was chosen from 98 random images of the *HPRT1* control. We noticed that for all plates there was a wide range of thresholds where the number of spots detected per image did not change significantly, as reported previously³. The threshold is chosen at the end of this plateau to increase the stringency

of the spot detection compared to granular background that does not indicate mRNA particles (**Supplementary Fig. 7c**). Note this did not affect the mean number of spots detected per cell (**Supplementary Fig. 7f**). As an additional safety measure against dim images and false-positive spots, at least one pixel within a spot required to have an intensity which would be slightly above the permitted range of the minimal intensity of an image (**Supplementary Fig. 7b**). Note that this safety measure would not generally impact the spot detection (not shown). Such a global approach circumvents the manual threshold selection for individual images described previously³. Also it increased the window of the signal between the expressed genes and negative controls since threshold detection algorithms without plate-wide limits would have a large number of false-positive spots that do not represent mRNA spots.

Spot bias correction

When we calculated the probability of each pixel in our field of view to be a the centroid of a detected spot we realized that there was a small bias ($\pm 2\%$) towards detecting spots located at the center of our field of view. Such bias is likely to arise from the large chip size of the sCMOS camera, allowing the lens curvature to have an effect (**Supplementary Fig. 7d**). The bias was corrected by locally applying different thresholds. In order to estimate the right threshold value per pixel in our field of view we applied 20-40 thresholds close the to the reference threshold for each individual plate before the main spot detection. For each threshold the amount of spots at a given position was determined and values represented as images, which were then smoothened using a Gaussian filter of 6.5um. Local scaling factors for the global threshold were determined in such a way that the local average would correspond to

the mean number of spots expected from the mean at the global reference threshold. These scaling factors were again smoothed using a Gaussian of 4.9 μ m before applying them to detect spots in order to avoid over-fitting.

Gaussian deblending of crowded spots

While the spot detection resolved most spots, it failed to properly resolve individual spots of highly abundant RNA species, such as mRNA of *ACTB*. To increase the dynamic range, intensity-based deblending was done to separate local peaks. We achieved this using a custom-made fast implementation in MATLAB of the deblending function from the astrophysics software SourceExtractor⁴. Instead of using a logarithmic scale on intensities (as is customary in astrophysics), we used a linear scale for deblending since the RNA spots tend to have similar intensities. (**Supplementary Fig. 7e**).

Supplementary Note 4: Calculation of fraction of cells with spot counts above background and mean spot per cell correction.

The fraction of cells with spot counts above the *dapB* control s_i was calculated using sampled distributions of each gene replicate normalized by the total number of cells of a given sample. For every given expression level i , when i is at least 1 spot per cell, s_i was defined as $s_i = d_i H(d_i)$, where d_i is the difference between the observed fraction of cells g_i for a given cell sample of a gene replicate and that of its corresponding *dapB* control c_i , and $H(d_i)$ is a Heaviside step function whose value is 0 for negative d_i and 1 for positive d_i . The total fraction of cells above the negative control S_g was then given by $S_g = \sum_{i=1}^N s_i$, where N is the maximum number of spots in one cell observed in both

replica experiments. The fraction of cell s_0 that did not show expression above *dapB*, i.e when $i = 0$, was then given by $s_0 = 1 - S_g$. The corrected mean expression μ_g was then obtained by $\mu_g = \frac{\sum_{i=0}^N s_i M}{M}$, where M is the number of sampled cells for that gene replicate. Calculation of S_g and μ_g was bootstrapped 100 times by subsampling one third of the cells for every gene replicate and the *dapB* control cells and mean $\overline{S_g}$ and $\overline{\mu_g}$ obtained for every replicate gene. The fraction above negative control expected by random was obtained for every multi-well plate in the screen by 10^4 bootstrapped comparisons of two subsampled populations of the *dapB* control. The bootstrapped distribution obtained for every replicate gene was then compared to the distributions of fractions expected by random using the Mann-Whitney-Wilcoxon test and p values obtained corrected for multiple testing using the Holm-Bonferroni method. To identify genes with fractions of cells above background, we set a conservative significance value of $p=10^{-4}$. Note that *dapB* control wells did not reach significance even when a less conservative significance level ($p=0.05$) was applied.

Supplementary Note 5: Identification of cellular patterns of mRNA localization.

Feature generation

Primary spot features were obtained by the custom “MeasureLocalizationOfSpots.m” CellProfiler⁵ module (see **Fig. 2** and module help on **Supplementary Software**). Of 18 primary features calculated only 16 were used for analysis:

1. Closest distance to membrane

2. Distance to cell centroid
3. Distance to nuclear centroid
4. Distance to cell outline projecting nuclear centroid
5. Mean distance to remaining spots
6. S.d of distance to remaining spots
- 7-12. Radius in pixels to include 5%, 10%, 15%, 25% 50% or 75% of all remaining spots in a cell. 13-16. Number of spots within 20, 40, 80 and 120 pixels from a given spot centre. Note that the feature describing the identity of the closest membrane was left out because it is more likely to describe the location of the cell in a population rather than the localization patterns of spots as cells with many neighbours are more likely to have spots close to a cell-to-cell boundary. For every cell the mean and s.d. of all above features were calculated and normalized to the square root of the area of a given cell, thus generating 32 cellular features.

Selection of data and genes for analysis

As mentioned in the main text, genes were considered for analysis if the comparison of the expression level obtained by RNA-seq and the mean number of spots per cell were consistent (including mitochondria genes) and if the mean expression level was 10 or more spots per cell (442 genes). To only take into account robust features, cells that had less than 10 spots per cell were also discarded from the analysis.

Cell sampling and hierarchical clustering of single cells

To make sure that sampled cells represented the full range of the 32-dimensional feature space and accounted for potential overrepresentation in mRNA localization due to more abundant cell shapes or sizes we devised a two-step sampling procedure. First, $\sim 7 \times 10^4$ cells were randomly sampled from the initial pool of cells and the pairwise Euclidean distance to a randomly sampled 20% of the 7×10^4 cells was computed in the 32-dimensional feature space after z-score normalization. The number of neighbours of every cell was then calculated. We chose a distance for the definition of a neighbour so that most cells (99.9%) had at least one neighbour. We then grouped cells in 200 bins of similar number of neighbours and randomly sampled from each bin ~ 50 cells. This resulted in the sampling of $\sim 10,000$ cells per cell clustering run. Hierarchical clustering of 3,960 cell-samplings was done using a Euclidian distance space and Ward's linkage method. We chose a final number of clusters so as to maximize the adjusted rand index⁶ of two different classifications of the same cells. This was done as follows: the results of the clustering of a given sample of cells were taken to be the classification *A* for that given sample set. Classification *B* was obtained by allocating to every cell in the given sampling set the cluster ID of the closest neighbouring cell from a second non-redundant sampling set. The adjusted rand index was then calculated for classifications *A* and *B*. On average $\sim 3,960$ pairwise comparisons had the maximum rand index when five clusters were assumed.

Definition of main localization types, computation of cellular mRNA localization phenotypes and description of gene localization patterns

To classify clusters from different sampling sets to common localization types we computed the centroid for every obtained cluster and grouped them using hierarchical clustering into five mRNA localization types. To measure the mRNA localization phenotype of a cell we measured the distance to a number of randomly sampled cluster centroids and defined the localization type of the cell as that of its closest centroid. Such classifications were run 10,000 times for every cell and the phenotype of a cell defined as the vector of fractions of times a cell was defined as belonging to a particular localization type. Note that classification using seven centroids per iteration yielded the largest reproducibility between the two replicate experiments (**Supplementary Fig. 11g**). From the above analysis the localization pattern of a given gene can be described by the distributions of the classification fraction given by a population of cells for every localization type. We describe such distributions by computing the mean, variance, and the 3rd and 4th central moments. See <https://github.com/pelkmanslab/locpatterns> for example MATLAB code.

Supplementary Note 6: Preparation of multivariate transcript readouts for clustering and feature selection

Feature generation

Treatment of primary spatial features: Primary spot features were obtained by the custom “MeasureLocalizationOfSpots.m” CellProfiler module (**Fig. 2, Online Methods and Supplementary Software**). Measurements of individual RNA-spots were

normalized by z-scoring against random cytoplasmic pixels. The number of sampled pixels would be the same as the number of spots, which were originally identified in the given cells. Sampling was done with full randomization without pixel replacement. Sampling of pixels was iterated 100 times. Per iteration, the measurements of a single random pixel were used for the normalization of the measurements of a given observed RNA-spot. Normalization was done by assigning observed spots the corresponding z-scored value of the 100 pixel distribution.

Generation of cellular features: For primary spatial features, cellular features were obtained by the custom "MeasureChildren.m" CellProfiler module (**Supplementary Software**). For individual parent objects ("Cells"), measurements robust to NaN (Not a Number) were derived from the normalized primary features of its children. The statistics measured were mean, median, standard deviation (s.d.) and variance as well as third, fourth, fifth and sixth central moments. For individual parent objects ("Cells"), only children ("Spot") measurements that were not NaN, were considered for calculation of the central moments. If this was not possible for a given set of primary features, the cellular feature would be set to NaN. For spot count features, additional cellular features (besides the raw spot count) were obtained by log-transforming the raw spot count and by division of the raw spot count by the area of the cell (defined as pixels within 2D segmentation of nucleus and cytoplasm) and multiplication with the median area of all cells included in either replicate assay (**Supplementary Fig. 15b**).

Generation of well features: Well features were obtained using a custom Matlab script following CellProfiler analysis. Measurements robust to NaN were derived from the cellular features. In this case we computed the mean, median, and variance using Matlab's built-in NaN-robust implementations thereof. NaN-robust 3rd to 6th central moments were created by considering only cells, where the measurement was not NaN. If no such cell was present, the given well measurement would be NaN. Note that a single well would correspond to a single gene within one replicate assay. Well measurements, containing an infinite number, were substituted with NaN.

Selection of data and genes for analysis

Genes were considered for analysis if the comparison of the expression level obtained by RNA-seq and mean number of spots was consistent. This was true for genes, which were no outliers in the direct comparison of mean expression levels, and for mitochondrially encoded RNAs (13 genes). In addition, the mean number of spots per cell had to be more than 10 in both replicates (442 genes).

Starting sets

Starting sets were the initial groups of features prior to feature selection and further analysis. Individual starting sets could have overlapping features. ALL starting features included mean and variability per well features of spot count and spatial features and the mean and variability of the classification of localization patterns (**Supplementary Fig. 15b**).

Data pre-processing

Data pre-processing was done separately for each starting set. Features were discarded prior to feature selection, if in at least 10% of the genes of an individual assay they were NaNs (leaving 620 of 638 features of ALL starting features, **Supplementary Fig. 14b**). Remaining NaNs (561 of 548,080 datapoints of ALL starting features) were approximated using a weighted average of the corresponding feature of the five closest genes in the Euclidean space (using MATLAB's `knnimpute`).

Feature selection

Genes were randomly split into two equally sized non overlapping groups. The first group served as a training set for feature selection, whereas the second group served to estimate the reproducibility after feature selection (**Supplementary Fig. 14c**). Features of the training set were Z-scored separately for each replicate assay. The fraction of highly reproducible genes was determined as described in Online Methods. Individual features were removed separately. The feature in whose absence there would be the highest fraction of highly reproducible genes was chosen. In case that the absence of multiple features would yield the same highest fraction of reproducible genes, one of them was chosen randomly. Next, the chosen feature was removed. Using the same training set of genes, features were iteratively removed until no feature was left. Afterwards, the fraction of the genes of the training set that was highly reproducible after each round of feature elimination was determined. These fractions were smoothened with a sliding average over 4 consecutive rounds of feature elimination after padding the earliest and latest values by replication. The set of features with the highest value was selected. In case that multiple sets would share the highest value, the

smaller feature set was selected. The feature selection was done 60 times with different randomization events and thus distinct training sets.

Union of replicate assays for clustering and overlap with STRING 9.0

Features of individual replicate assays were normalized independently by z-scoring and then averaged over both replicate assays.

Supplementary Tables

Supplementary Table 1: Spot count per cell upon gene knock-down by RNA interference

Probes	siRNA	Wells, bDNA FISH			Number	Single cells, bDNA FISH				Fold change		bDNA FISH		qPCR	
		Mean	Stdev			Mean	Stdev	Number	CV	Mean	Stdev	Mean	Stdev	Mean	Stdev
TFRC	none	187.937	6.878		4	187.733	72.460	13088	0.037						
TFRC	scrambled	183.485	6.526		4	183.426	75.609	12631	0.036						
TFRC	TFRC	37.895	8.322		4	38.418	22.868	12168	0.220	0.207	0.045	0.181	0.009		
MYC	none	203.013	8.026		4	203.100	86.795	10900	0.040						
MYC	scrambled	161.530	16.849		4	162.352	75.063	10393	0.104						
MYC	MYC	26.099	3.974		4	26.271	33.690	5791	0.152	0.162	0.025	0.176	0.018		
ERBB2	none	23.413	0.468		4	23.403	11.087	13082	0.020						
ERBB2	scrambled	17.963	1.071		4	18.037	9.525	12230	0.060						
ERBB2	ERBB2	4.199	0.340		4	4.197	3.651	7253	0.081	0.234	0.019	0.199	0.006		
dapB	none	0.484	0.053		6	0.482	1.002	21094	0.109						
none	none	0.304	0.044		6	0.303	0.992	19546	0.145						

bDNA FISH and qPCR Mean Fold Change:
number of qPCR runs
CV:

gene specific siRNA / scrambled
3 all conditions
Coefficient of variation

Supplementary Table 2: Targeted genes in the library and their Affymetrix product ID

Affymetrix Catalog#	Gene Symbol	Entrez ID
VA1-10006	IL6	3569
VA1-10007	IL8	3576
VA1-10009	ZNF189	7743
VA1-10010	BLM	641
VA1-10011	GLI	2735
VA1-10015	PDZK1	5174
VA1-10016	TFF1	7031
VA1-10018	G6PC	2538
VA1-10022	KRT19	3880
VA1-10025	PLAUR	5329
VA1-10027	SERPINE1	5054
VA1-10029	ERCC6L	54821
VA1-10032	SIDT1	54847
VA1-10040	CDKN2A	1029
VA1-10046	Stat3	6774
VA1-10047	ATG12	9140
VA1-10048	CGB	1082
VA1-10050	ESCO2	157570
VA1-10051	HEMGN	55363
VA1-10052	CA1	759
VA1-10053	HBG2	3048
VA1-10054	HBE1	3046
VA1-10061	PVT1	5820
VA1-10069	SOD1	6647
VA1-10070	COL1A1	1277
VA1-10073	UGCG	7357
VA1-10074	LMO2	4005
VA1-10076	ETS2	2114
VA1-10078	CA2	760
VA1-10079	HLA-G	3135
VA1-10080	HBG1	3047
VA1-10085	IL13RA2	3598
VA1-10096	ABL1	25
VA1-10099	INS	3630
VA1-10113	MMP2	4313
VA1-10114	MMP1	4312
VA1-10119	GAPDH	2597
VA1-10122	IL17F	112744
VA1-10123	IL17	3605
VA1-10124	HMOX1	3162
VA1-10126	DDIT3	1649
VA1-10128	BTG2	7832
VA1-10130	LGALS3	3958
VA1-10134	fn1	2335
VA1-10148	PPIB	5479
VA1-10151	KRT7	3855
VA1-10156	PPARG	5468
VA1-10159	BMP6	654
VA1-10160	ABCA1	19
VA1-10167	HCRT	3060
VA1-10168	igfbp3	3486

Affymetrix Catalog#	Gene Symbol	Entrez ID
VA1-12073	prkcq	5588
VA1-12074	PRKCD	5580
VA1-12075	prkaa2	5563
VA1-12076	PRKAA1	5562
VA1-12077	nfe2l2	4780
VA1-12078	RPL10P15	6134
VA1-12079	MCM7	4176
VA1-12080	SMAD7	4092
VA1-12081	SMAD3	4088
VA1-12082	SMAD2	4087
VA1-12083	IDH1	3417
VA1-12084	STAM2	10254
VA1-12085	STX6	10228
VA1-12086	flot1	10211
VA1-12087	PSME3	10197
VA1-12088	Actr2	10097
VA1-12089	ACTR3	10096
VA1-12090	VAMP7	6845
VA1-12091	SOS1	6654
VA1-12092	DLL1	28514
VA1-12093	rheb	6009
VA1-12094	PTK2	5747
VA1-12095	SMAD6	4091
VA1-12096	lamp1	3916
VA1-12097	KARS	3735
VA1-12098	Akt3	10000
VA1-12099	TP73	7161
VA1-12100	Src	6714
VA1-12101	Prkce	5581
VA1-12102	RAB8A	4218
VA1-12103	smad4	4089
VA1-12104	JUND	3727
VA1-12105	HSP90AA1	3320
VA1-12106	hadh	3033
VA1-12107	cttn	2017
VA1-12108	E2F1	1869
VA1-12109	cbl	867
VA1-12110	CAPN1	823
VA1-12111	akt1	207
VA1-12112	usp8	9101
VA1-12113	PLK1	5347
VA1-12114	pfn1	5216
VA1-12115	Oxa1l	5018
VA1-12116	myo6	4646
VA1-12117	MYO1E	4643
VA1-12118	ABCC1	4363
VA1-12119	Hdac1	3065
VA1-12120	Eif4g1	1981
VA1-12121	DNM2	1785
VA1-12122	cdkn2b	1030
VA1-12123	CLOCK	9575

VA1-10169	HK2	3099
VA1-10174	XBP1	7494
VA1-10179	RELN	5649
VA1-10184	BDNF	627
VA1-10185	HIF1A	3091
VA1-10188	18S	
VA1-10189	BCL2	596
VA1-10196	CYP3A4	1576
VA1-10197	Apob	338
VA1-10201	ENG	2022
VA1-10203	UBC	7316
VA1-10204	FOLH1	2346
VA1-10205	cav1	857
VA1-10206	BCL2L10	10017
VA1-10214	ELAVL1	1994
VA1-10215	ADRA1A	148
VA1-10217	SP7	121340
VA1-10219	KIF11	3832
VA1-10220	FGF21	26291
VA1-10243	AHSA1	10598
VA1-10253	GFAP	2670
VA1-10263	CXCR4	7852
VA1-10265	TNFRSF12A	51330
VA1-10266	KLF2	10365
VA1-10267	TSLP	85480
VA1-10268	CD207	50489
VA1-10269	LAMP3	27074
VA1-10270	CD80	941
VA1-10271	DD3-PCA3	50652
VA1-10272	dapB	944762
VA1-10275	CSF2	1437
VA1-10300	ACTA2	59
VA1-10301	SST	6750
VA1-10302	GLP1R	2740
VA1-10303	CALCR	799
VA1-10304	CALCA	796
VA1-10313	ITGAM	3684
VA1-10322	PCSK9	255738
VA1-10324	STK33	65975
VA1-10325	EZH2	2146
VA1-10326	MTOR	2475
VA1-10327	MYC	4609
VA1-10343	GDF2	2658
VA1-10344	ANLN	54443
VA1-10345	BRRN1	23397
VA1-10346	UBE2C	11065
VA1-10347	KIF2C	11004
VA1-10348	ACP5	54
VA1-10349	APOA1	335
VA1-10351	ACTB	60
VA1-10353	MACC1	346389
VA1-10357	TGFA	7039
VA1-10358	DTR	1839
VA1-10359	PTGS2	5743
VA1-10360	CD44	960
VA1-10361	HGF	3082
VA1-10362	SPP1	6696
VA1-10363	MET	4233

VA1-12124	CLTC	1213
VA1-12125	atg5	9474
VA1-12126	Abcg2	9429
VA1-12127	Pex16	9409
VA1-12128	zfyve9	9372
VA1-12129	etf1	2107
VA1-12130	HGS	9146
VA1-12131	RABEP1	9135
VA1-12132	Rab11a	8766
VA1-12133	SDPR	8436
VA1-12134	pabpn1	8106
VA1-12135	RAB7A	7879
VA1-12136	tgfbfr1	7046
VA1-12137	TAF1	6872
VA1-12138	STX1A	6804
VA1-12139	rab5c	5878
VA1-12140	rab4a	5867
VA1-12141	PPP2R2B	5521
VA1-12142	NOTCH4	4855
VA1-12143	mcm2	4171
VA1-12144	NRG1	3084
VA1-12145	GOLGA2	2801
VA1-12146	FLOT2	2319
VA1-12147	FDFT1	2222
VA1-12148	endog	2021
VA1-12149	DVL3	1857
VA1-12150	DVL2	1856
VA1-12151	DVL1	1855
VA1-12152	DNM1	1759
VA1-12153	CREBBP	1387
VA1-12154	CREB1	1385
VA1-12155	CDH1	999
VA1-12156	Casp3	836
VA1-12157	BAX	581
VA1-12158	Bad	572
VA1-12159	arrb2	409
VA1-12160	bin1	274
VA1-12161	ACVR1B	91
VA1-12162	RAB11B	9230
VA1-12163	AIFM1	9131
VA1-12164	wars	7453
VA1-12165	rab5a	5868
VA1-12166	psmb8	5696
VA1-12167	HSPA9	3313
VA1-12168	FEN1	2237
VA1-12169	FASN	2194
VA1-12170	EIF4EBP1	1978
VA1-12171	Echs1	1892
VA1-12172	AP2S1	1175
VA1-12173	AP2M1	1173
VA1-12174	CDKN1B	1027
VA1-12175	ARRB1	408
VA1-12176	NFKB1	4790
VA1-12177	SH2D2A	9047
VA1-12178	WASL	8976
VA1-12179	btrc	8945
VA1-12180	SYNJ2	8871
VA1-12181	SYNJ1	8867

VA1-10364	HCRT2	3062
VA1-10366	VCP	7415
VA1-10367	TTK	7272
VA1-10368	NUP98	4928
VA1-10369	CCNB1	891
VA1-10373	KRT5	3852
VA1-10374	GAL	51083
VA1-10376	ADIPOR2	79602
VA1-10377	ADIPOR1	51094
VA1-10381	KLK3	354
VA1-10388	AXIN2	8313
VA1-10395	PDGFRB	5159
VA1-10396	NRP1	8829
VA1-10400	PROM1	8842
VA1-10417	RPS6	6194
VA1-10418	EEF1A1	1915
VA1-10437	ACTG2	72
VA1-10441	CDX2	1045
VA1-10449	RUNX2	860
VA1-10450	COL2A1	1280
VA1-10451	AGC1	176
VA1-10452	SOX9	6662
VA1-10453	NCOA3	8202
VA1-10456	OTUD5	55593
VA1-10478	DKK1	22943
VA1-10479	BMP2	650
VA1-10481	TNF	7124
VA1-10483	FPR2	2358
VA1-10484	Pak1	5058
VA1-10511	ZNF174	7727
VA1-10512	HEXIM1	10614
VA1-10515	IL17RB	55540
VA1-10516	IL25	64806
VA1-10519	id1	3397
VA1-10521	FGFR3	2261
VA1-10531	ADRA1D	146
VA1-10532	adra1b	147
VA1-10537	ND5	4540
VA1-10544	S100a4	6275
VA1-10545	VEGFA	7422
VA1-10581	IL23A	51561
VA1-10586	LGR6	59352
VA1-10587	GPR49	8549
VA1-10603	PMEP1	56937
VA1-10604	OR51E2	81285
VA1-10608	NKX3-1	4824
VA1-10610	ERG	2078
VA1-10611	B2M	567
VA1-10612	SPINK1	6690
VA1-10616	Ar	367
VA1-10624	ROS1	6098
VA1-10625	ALK	238
VA1-10628	CMKOR1	57007
VA1-10629	GPR101	83550
VA1-10630	GPR123	84435
VA1-10642	GLI2	2736
VA1-10647	KRT6A	3853
VA1-10649	P2RY11	5032

VA1-12182	FBP2	8789
VA1-12183	snap23	8773
VA1-12184	TNFSF10	8743
VA1-12185	SNX4	8723
VA1-12186	BECN1	8678
VA1-12187	numb	8650
VA1-12188	Yars	8565
VA1-12189	AP3B1	8546
VA1-12190	pex3	8504
VA1-12191	dyrk3	8444
VA1-12192	Soat2	8435
VA1-12193	EEA1	8411
VA1-12194	PLA2G10	8399
VA1-12195	pla2g6	8398
VA1-12196	Pip5k1b	8395
VA1-12197	fzd1	8321
VA1-12198	stam	8027
VA1-12199	fzd5	7855
VA1-12200	xpo1	7514
VA1-12201	WNT9A	7483
VA1-12202	WNT5A	7474
VA1-12203	EZR	7430
VA1-12204	vcl	7414
VA1-12205	SUMO1	7341
VA1-12206	UBE2I	7329
VA1-12207	Tgfbr2	7048
VA1-12208	TFRC	7037
VA1-12209	STX5	6811
VA1-12210	SRPR	6734
VA1-12211	srp54	6729
VA1-12212	SRP19	6728
VA1-12213	SRP14	6727
VA1-12214	SOX2	6657
VA1-12215	soat1	6646
VA1-12216	SNAP25	6616
VA1-12217	SHC1	6464
VA1-12218	sdhD	6392
VA1-12219	SdhB	6390
VA1-12220	Sort1	6272
VA1-12221	RPA1	6117
VA1-12222	rfc1	5981
VA1-12223	RASGRF1	5923
VA1-12224	RAF1	5894
VA1-12225	RAB13	5872
VA1-12226	rab5b	5869
VA1-12227	PXN	5829
VA1-12228	pex19	5824
VA1-12229	Ptx3	5806
VA1-12230	PTPN6	5777
VA1-12231	ptpn1	5770
VA1-12232	PSMD4	5710
VA1-12233	psmb9	5698
VA1-12234	Psma6	5687
VA1-12235	MAP2K1	5604
VA1-12236	mapk8	5599
VA1-12237	MAPK1	5594
VA1-12238	Prkcz	5590
VA1-12239	PRKCI	5584

VA1-10656	TRIB2	28951
VA1-10721	TMPRSS2	7113
VA1-10723	AMACR	23600
VA1-10725	PCA3	50652
VA1-10733	HOXB9	3219
VA1-10745	GHSR	2693
VA1-10746	mapt	4137
VA1-10747	GPR83	10888
VA1-10748	GPR35	2859
VA1-10749	SSTR3	6753
VA1-10750	GPER	2852
VA1-10790	LRRC31	79782
VA1-10791	FLT4	2324
VA1-10816	BCL2L11	10018
VA1-10840	IL10	3586
VA1-10846	EIF4E	1977
VA1-10847	L1TD1	54596
VA1-10851	PLAU	5328
VA1-10852	FSTL1	11167
VA1-10870	COL4A1	1282
VA1-10871	SEC31B	25956
VA1-10872	PECAM1	5175
VA1-10879	ENAH	55740
VA1-10880	D4S234E	27065
VA1-10881	DEF6	50619
VA1-10882	PCDHB15	56121
VA1-10885	TRPV1	7442
VA1-10886	TRPM2	7226
VA1-10904	KRT17	3872
VA1-10938	CALCRL	10203
VA1-10939	SCN9A	6335
VA1-10943	Notch2	4853
VA1-10957	MAGEA2	4101
VA1-10961	TP73L	8626
VA1-11000	LRG1	116844
VA1-11001	EREG	2069
VA1-11010	FAM123A	219287
VA1-11011	FAM123C	205147
VA1-11020	GPC3	2719
VA1-11030	MCM6	4175
VA1-11031	ARPC2	10109
VA1-11032	CLEC3B	7123
VA1-11033	MKI67	4288
VA1-11035	CXCL12	6387
VA1-11041	CCL2	6347
VA1-11042	CCR2	729230
VA1-11043	NOS2	4843
VA1-11055	RAVER2	55225
VA1-11056	TGM4	7047
VA1-11057	SEMG2	6407
VA1-11070	HRH3	11255
VA1-11072	GAD1	2571
VA1-11073	HCRTR1	3061
VA1-11084	HSPA5	3309
VA1-11095	KRT14	3861
VA1-11096	KRT10	3858
VA1-11103	PPIA	5478
VA1-11104	OLFM4	10562

VA1-12240	PPP2CA	5515
VA1-12241	POLR2G	5436
VA1-12242	POLR2E	5434
VA1-12243	PML	5371
VA1-12244	PLD2	5338
VA1-12245	PLD1	5337
VA1-12246	PIK3CG	5294
VA1-12247	SERPINB5	5268
VA1-12248	PGF	5228
VA1-12249	pfn2	5217
VA1-12250	pfkp	5214
VA1-12251	PFKL	5211
VA1-12252	Per1	5187
VA1-12253	PDPK1	5170
VA1-12254	Pak2	5062
VA1-12255	PABPC1	26986
VA1-12256	OCLN	1.01E+08
VA1-12257	MSI1	4440
VA1-12258	mmp7	4316
VA1-12259	MDM4	4194
VA1-12260	mdm2	4193
VA1-12261	MCM3	4172
VA1-12262	ldhb	3945
VA1-12263	lamp2	3920
VA1-12264	KDR	3791
VA1-12265	Jun	3725
VA1-12266	ITGA5	3678
VA1-12267	ITGA3	3675
VA1-12268	ACO1	48
VA1-12269	IDH2	3418
VA1-12270	HSPD1	3329
VA1-12271	HMGB1	3146
VA1-12272	GTF2F1	2962
VA1-12273	GSK3B	2932
VA1-12274	GRB2	2885
VA1-12275	Gars	2617
VA1-12276	FGF2	2247
VA1-12277	EPS15	2060
VA1-12278	eno2	2026
VA1-12279	Egf	1950
VA1-12280	DCN	1634
VA1-12281	CTSD	1509
VA1-12282	CTSB	1508
VA1-12283	csnk2a2	1459
VA1-12284	CSNK2A1	1457
VA1-12285	CLTA	1211
VA1-12286	Cel	1056
VA1-12287	Cdc42	998
VA1-12288	Cdk1	983
VA1-12289	CAPN2	824
VA1-12290	ATP5B	506
VA1-12291	Rhoa	387
VA1-12292	Arf6	382
VA1-12293	Arf1	375
VA1-12294	Apex1	328
VA1-12295	ampH	273
VA1-12296	akt2	208
VA1-12297	parp1	142

VA1-11109	Mmp9	4318
VA1-11113	KIF20A	10112
VA1-11115	DPY19L1	23333
VA1-11121	PIK3CA	5290
VA1-11122	SPARC	6678
VA1-11123	PCGEM1	64002
VA1-11124	HPRT1	3251
VA1-11127	TBP	6908
VA1-11131	BDKRB1	623
VA1-11133	pcnA	5111
VA1-11134	BCL6	604
VA1-11135	BRAF	673
VA1-11137	BIRC5	332
VA1-11144	MYF5	4617
VA1-11152	tp53	7157
VA1-11167	GPR142	350383
VA1-11168	GCG	2641
VA1-11180	MTNR1B	4544
VA1-11181	MTNR1A	4543
VA1-11187	F2RL1	2150
VA1-11211	RPLP0	6175
VA1-11212	GUSB	2990
VA1-11257	FEV	54738
VA1-11265	P2RX7	5027
VA1-11269	MSLN	10232
VA1-11274	GPR40	2864
VA1-11276	RALA	5898
VA1-11277	RALB	5899
VA1-11281	IFNB1	3456
VA1-11286	KRT15	3866
VA1-11301	IL4	3565
VA1-11313	MYOD1	4654
VA1-11317	MALAT1	378938
VA1-11327	ISL1	3670
VA1-11387	IFI27	3429
VA1-11388	USP18	11274
VA1-11389	STAT1	6772
VA1-11391	SERPINA1	5265
VA1-11392	C11orf82	220042
VA1-11393	KRT16	3868
VA1-11397	HNF4	3172
VA1-11414	HOTAIRM1	1.01E+08
VA1-11415	HSD11B1	3290
VA1-11416	GALR2	8811
VA1-11417	KCNQ2	3785
VA1-11418	KCNQ3	3786
VA1-11425	PTP4A3	11156
VA1-11432	CHAT	1103
VA1-11454	CD68	968
VA1-11465	FAM123B	139285
VA1-11466	TG	7038
VA1-11501	ERCC1	2067
VA1-11502	TYMS	7298
VA1-11503	ESR1	2099
VA1-11504	FOS	2353
VA1-11506	KIT	3815
VA1-11548	APP	351
VA1-11552	FLI1	2313

VA1-12298	INPP1	3628
VA1-12299	gtf2i	2969
VA1-12300	Gtf2b	2959
VA1-12301	EP300	2033
VA1-12302	CELSR2	1952
VA1-12303	DYRK1A	1859
VA1-12304	DKC1	1736
VA1-12305	DAB2	1601
VA1-12306	Ctbp1	1487
VA1-12307	CSNK2B	1460
VA1-12308	CLDN3	1365
VA1-12309	CLDN4	1364
VA1-12310	AP2B1	163
VA1-12311	CAV3	859
VA1-12312	CAV2	858
VA1-12313	casp10	843
VA1-12314	Casp9	842
VA1-12315	CASP8	841
VA1-12316	CASP6	839
VA1-12317	ARHGAP5	394
VA1-12318	XIAP	331
VA1-12319	BIRC3	330
VA1-12320	Acacb	32
VA1-12321	TF	7018
VA1-12322	rps23	6228
VA1-12323	RPS19P3	6223
VA1-12324	murC	347273
VA1-12325	POLR2B	5431
VA1-12326	POLR2A	5430
VA1-12327	ABCB1	5243
VA1-12328	Igf2r	3482
VA1-12329	HMGCR	3156
VA1-12330	FOLR1	2348
VA1-12331	fgf1	2246
VA1-12332	CYP27B1	1594
VA1-12333	ALDH2	217
VA1-12334	ache	43
VA1-12335	TGFB1	7040
VA1-12336	FASLG	356
VA1-12337	SLC11A1	6556
VA1-12338	vwf	7450
VA1-12339	vhl	7428
VA1-12340	TSC2	7249
VA1-12341	Ldlr	3949
VA1-12342	Fbp1	2203
VA1-12343	CFTR	1080
VA1-12344	STK11	6794
VA1-12345	NOTCH3	4854
VA1-12346	ercc2	2068
VA1-12347	CDKN1A	1026
VA1-12348	TSC1	7248
VA1-12349	RB1	5925
VA1-12350	PPT1	5538
VA1-12351	PKLR	5313
VA1-12352	PGK1	5230
VA1-12353	PFKM	5213
VA1-12354	Pdha1	5160
VA1-12355	lipA	3988

VA1-11553	CD274	29126
VA1-11566	CLU	1191
VA1-11584	KRT18	3875
VA1-11586	KRT8	3856
VA1-11600	Kras	3845
VA1-11601	CTNNB1	1499
VA1-11608	IL1R1	3554
VA1-11609	IL1B	3553
VA1-11610	Il1a	3552
VA1-11611	IL6R	3570
VA1-11612	P2RY14	9934
VA1-11634	ISG15	9636
VA1-11635	IFNG	3458
VA1-11638	IL29	282618
VA1-11654	XIST	7503
VA1-11656	VPS13A	23230
VA1-11697	TTBK1	84630
VA1-11698	TTBK2	146057
VA1-11699	GABRB3	2562
VA1-11700	GABRB2	2561
VA1-11701	GABRB1	2560
VA1-11702	GABRA2	2555
VA1-11703	GABRA1	2554
VA1-11704	NOS1	4842
VA1-11705	NOS3	4846
VA1-11709	FABP1	2168
VA1-11714	ERBB2	2064
VA1-11715	ERBB4	2066
VA1-11716	ERBB3	2065
VA1-11724	PSAT1	29968
VA1-11727	COPZ2	51226
VA1-11728	COPG2	26958
VA1-11736	egfr	1956
VA1-11737	KRT23	25984
VA1-11738	MITF	4286
VA1-11740	pten	5728
VA1-11742	TERT	7015
VA1-11753	ETV4	2118
VA1-11754	ETV5	2119
VA1-11755	PTPRC	5788
VA1-11756	CD34	947
VA1-11757	KCNK18	338567
VA1-11760	pou5f1	5460
VA1-11763	TRPA1	8989
VA1-11764	GABRA3	2556
VA1-11790	ETV1	2115
VA1-11808	ASPM	259266
VA1-11836	FZD10	11211
VA1-11837	MPO	4353
VA1-11840	CX3CR1	1524
VA1-11870	IGHG4	3503
VA1-11872	SSTR2	6752
VA1-11879	TUG1	55000
VA1-11891	rac1	5879
VA1-11893	HPSE	10855
VA1-11894	RBMY1A1	5940
VA1-11905	JAK2	3717
VA1-11906	SNORD3A	780851

VA1-12356	LCAT	3931
VA1-12357	JAK3	3718
VA1-12358	ITGB3	3690
VA1-12359	INSR	3643
VA1-12360	ICAM1	3383
VA1-12361	hadhb	3032
VA1-12362	HADHA	3030
VA1-12363	Gsn	2934
VA1-12364	NR3C1	2908
VA1-12365	GPI	2821
VA1-12366	Gla	2717
VA1-12367	GBA	2629
VA1-12368	FH	2271
VA1-12369	ercc3	2071
VA1-12370	taz	6901
VA1-12371	dld	1738
VA1-12372	CD36	948
VA1-12373	C3	718
VA1-12374	Brca2	675
VA1-12375	ATM	472
VA1-12376	Fas	355
VA1-12377	APC	324
VA1-12378	Adrb3	155
VA1-12379	ADRB2	154
VA1-12380	Psen1	5663
VA1-12381	ND3	4537
VA1-12382	ND2	4536
VA1-12383	ND1	4535
VA1-12384	CYTB	4519
VA1-12385	COX3	4514
VA1-12386	ATP6	4508
VA1-12387	COX2	4513
VA1-12388	COX1	4512
VA1-12389	ND6	4541
VA1-12390	ND4L	4539
VA1-12391	ND4	4538
VA1-12392	ATP8	4509
VA1-12393	CD209	30835
VA1-12394	Ran	5901
VA1-12395	LdhA	3939
VA1-12396	SEPP1	6414
VA1-12397	PIP5K1A	8394
VA1-12398	OGG1	4968
VA1-12399	M6PR	4074
VA1-12400	rpsA	3921
VA1-12401	FYN	2534
VA1-12402	eno1	2023
VA1-12403	Pfn3	345456
VA1-12404	rps3	6188
VA1-12405	trnL1	4567
VA1-12406	PKP4	8502
VA1-12407	TJP1	7082
VA1-12408	FOXO1A	2308
VA1-12410	HOPX	84525
VA1-12416	DLL4	54567
VA1-12423	HOXB13	10481
VA1-12424	GPR44	11251
VA1-12425	HOXB7	3217

VA1-11907	RNU2-1	6066
VA1-11908	ASCL1	429
VA1-11909	SCN3A	6328
VA1-11910	SCN10A	6336
VA1-11911	NTRK2	4915
VA1-11912	NTRK3	4916
VA1-11913	NTRK1	4914
VA1-11922	PCSK5	5125
VA1-11928	NGEF	25791
VA1-11931	PITX3	5309
VA1-11932	NR4A2	4929
VA1-11944	ACCN2	41
VA1-11945	CALCB	797
VA1-11946	ADRA2C	152
VA1-11947	ADRA2B	151
VA1-11948	ADRA2A	150
VA1-11949	ACCN3	9311
VA1-11961	S100B	6285
VA1-11963	AQP4	361
VA1-11964	WNT2	7472
VA1-11965	prickle2	166336
VA1-11966	ACACA	31
VA1-11967	Pik3r1	5295
VA1-11968	CHMP4B	128866
VA1-11969	PRICKLE1	144165
VA1-11970	CHMP7	91782
VA1-11971	PRKCDBP	112464
VA1-11972	Ehd4	30844
VA1-11973	VANGL1	81839
VA1-11974	FCHO2	115548
VA1-11975	TLR4	7099
VA1-11976	sp1	6667
VA1-11977	Itgb1	3688
VA1-11978	CCND1	595
VA1-11979	Fam125b	89853
VA1-11980	Pard6b	84612
VA1-11981	vps25	84313
VA1-11982	Sh3kbp1	30011
VA1-11983	Tcf7l2	6934
VA1-11984	Map2k2	5605
VA1-11985	rufy1	80230
VA1-11986	RAB11FIP1	80223
VA1-11987	NANOG	79923
VA1-11988	arhgap10	79658
VA1-11989	CHMP6	79643
VA1-11990	pla2g4a	5321
VA1-11991	CTBP2	1488
VA1-11992	MICAL1	64780
VA1-11993	SMURF2	64750
VA1-11994	ELOVL4	6785
VA1-11995	ZFYVE20	64145
VA1-11996	relA	5970
VA1-11997	CLDN1	9076
VA1-11998	VPS18	57617
VA1-11999	RPTOR	57521
VA1-12000	RAB22A	57403
VA1-12001	SMURF1	57154
VA1-12002	CHMP1B	57132

VA1-12427	IGLC	
VA1-12428	DNTT	1791
VA1-12429	CTCFL	140690
VA1-12430	CTCF	10664
VA1-12435	PAF1	54623
VA1-12445	TCN1	6947
VA1-12450	IGF2	3481
VA1-12451	EPCAM	4072
VA1-12452	CREB3L1	90993
VA1-12482	THY1	7070
VA1-12522	JAG1	182
VA1-12523	PDPN	10630
VA1-12537	CLEC4C	170482
VA1-12538	TLR7	51284
VA1-12539	PTPRN	5798
VA1-12540	PME-1	51400
VA1-12542	CCR7	1236
VA1-12546	IL15	3600
VA1-12550	IL15RA	3601
VA1-12551	IL2RB	3560
VA1-12569	CXCL17	284340
VA1-12570	WT1	7490
VA1-12580	MGMT	4255
VA1-12582	FXN	2395
VA6-10620	GLS	2744
VA6-10768	RORC	6097
VA6-10774	HOXD4	3233
VA6-10914	PRR11	55771
VA6-11198	ID2	3398
VA6-11199	ZNF691	51058
VA6-11201	TCF4	6925
VA6-11203	MXI1	4601
VA6-11204	KLF7	8609
VA6-11205	KLF3	51274
VA6-11244	DUSP6	1848
VA6-11245	DUSP5	1847
VA6-11248	CDH2	1000
VA6-11271	ATOH1	474
VA6-11305	SLC5A5	6528
VA6-11322	WNT11	7481
VA6-11410	NES	10763
VA6-11424	CD79B	974
VA6-11427	QARS	5859
VA6-11428	NKX2-5	1482
VA6-11429	GATA4	2626
VA6-11434	EML4	27436
VA6-11439	WNT10A	80326
VA6-11453	PDCD1LG2	80380
VA6-11535	ENY2	56943
VA6-11536	TMEM141	85014
VA6-11537	STK3	6788
VA6-11538	HRSP12	10247
VA6-11539	HNRNPR	10236
VA6-11540	HOXA1	3198
VA6-11541	RECQL	5965
VA6-11542	MX2	4600
VA6-11543	CDC20	991
VA6-11544	PRIM2	5558

VA1-12003	CLDN2	9075
VA1-12004	Vangl2	57216
VA1-12005	Sh3glb2	56904
VA1-12006	Pard3	56288
VA1-12007	ddit4	54541
VA1-12008	Vps35	55737
VA1-12009	EPN3	55040
VA1-12010	OXSM	54995
VA1-12011	notch1	4851
VA1-12012	myh2	4620
VA1-12013	PARD6A	50855
VA1-12014	DLL3	10683
VA1-12015	polA1	5422
VA1-12016	RAB8B	51762
VA1-12017	Chmp5	51510
VA1-12018	SNX9	51429
VA1-12019	VPS28	51160
VA1-12020	RAB4B	53916
VA1-12021	HSD17B12	51144
VA1-12022	VPS36	51028
VA1-12023	SBDS	51119
VA1-12024	ldlrp1	26119
VA1-12025	DNM3	26052
VA1-12026	KANK2	25959
VA1-12027	RAB11FIP5	26056
VA1-12028	FCHO1	23149
VA1-12029	arhgap26	23092
VA1-12030	EPN2	22905
VA1-12031	Rab11fip2	22841
VA1-12032	SNAP91	9892
VA1-12033	DNAJC6	9829
VA1-12034	ZFYVE16	9765
VA1-12035	RAB11FIP3	9727
VA1-12036	Ehd2	30846
VA1-12037	ehd3	30845
VA1-12038	ZNRD1	30834
VA1-12039	ARFGAP3	26286
VA1-12040	Mcat	27349
VA1-12041	CHMP2A	27243
VA1-12042	snx5	27131
VA1-12043	celsr1	9620
VA1-12044	vamp2	6844
VA1-12045	AP2A1	160
VA1-12046	CHMP4A	29082
VA1-12047	Chmp2b	25978
VA1-12048	epn1	29924
VA1-12049	Ap2a2	161
VA1-12050	PTRF	284119
VA1-12051	FZD4	8322
VA1-12052	BRCA1	672
VA1-12053	GABARAP	11337
VA1-12054	ACOT7	11332
VA1-12055	vps45	11311
VA1-12056	snf8	11267
VA1-12057	CHEK2	11200
VA1-12058	pemt	10400
VA1-12059	PICALM	8301
VA1-12060	TRIO	7204

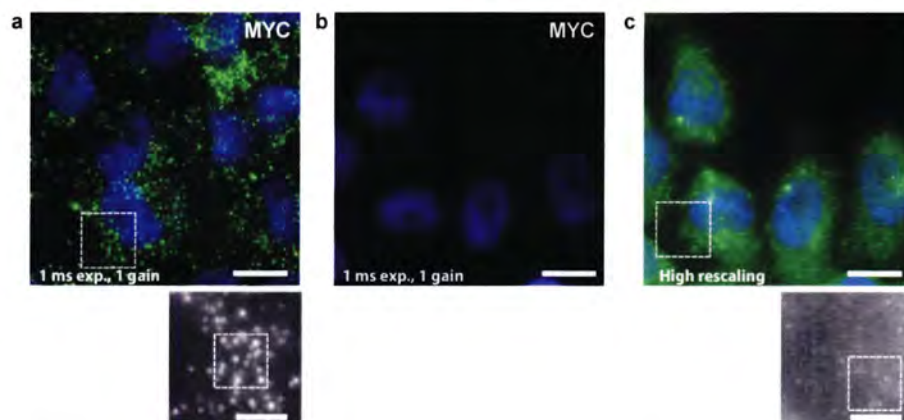
VA6-11568	MLANA	2315
VA6-11569	TYR	7299
VA6-11575	NUPR1	26471
VA6-11583	UGT2B15	7366
VA6-11598	TUBB3	10381
VA6-11599	MAP2	4133
VA6-11616	RMRP	6023
VA6-11761	PAX6	5080
VA6-11766	BMF	90427
VA6-11813	DBH	1621
VA6-11880	NRON	641373
VA6-11917	TLR9	54106
VA6-11927	GPRCSC	55890
VA6-11934	SLC6A3	6531
VA6-12440	SLC16A3	9123
VA6-12444	ZBED2	79413
VA6-12528	EPM2AIP1	9852
VA6-12565	FOXP3	50943
VA8-10340	GCLM	2730
VA8-10398	SMAD1	4086
VA8-10561	NKX6-1	4825
VA8-10621	GRIK2	2898
VA8-10777	HOXD3	3232
VA8-10856	NEAT1	283131
VA8-10857	STL	7955
VA8-10858	MEG3	55384
VA8-10916	COX11	1353
VA8-10918	CENPW	387103
VA8-11009	MAPK14	1432
VA8-11053	RLBP1	6017
VA8-11138	MAPKAPK2	9261
VA8-11615	PAX8AS	654433
VA8-12460	WDR78	79819
VA8-12461	GALNT10	55568
VA8-12462	ZNF547	284306
VA8-12463	GFRA1	2674
VA8-12464	GNAS	2778
VA8-12465	BAZ1B	9031
VA8-12467	CLSTN2	64084
VA8-12468	HCN1	348980
VA8-12469	CORIN	10699
VA8-12470	TRAF6	7189
VA8-12471	DMBT1	1755
VA8-12472	TNKS	8658
VA8-12473	GML	2765
VA8-12474	IFNAR1	3454
VA8-12475	IFNA2	3440
VA10-10295	CRTC2	200186
VA10-10296	CNR1	1268
VA10-10336	HSPA1A	3303
VA10-10339	SQSTM1	8878
VA10-10770	TSIX	9383
VA10-10772	TERF2IP	54386
VA10-10775	CCNDBP1	23582
VA10-10778	PDZK1IP1	10158
VA10-10779	NCOA4	8031
VA10-10781	GYPB	2994
VA10-10783	TBXA2R	6915

VA1-12061	CLTCL1	8218
VA1-12062	Cltb	1212
VA1-12063	SRP72	6731
VA1-12064	SOS2	6655
VA1-12065	rab31	11031
VA1-12066	EHD1	10938
VA1-12067	PNPLA6	10908
VA1-12068	exoc5	10640
VA1-12069	STAMBP	10617
VA1-12070	NXF1	10482
VA1-12071	XRCC1	7515
VA1-12072	Tsg101	7251

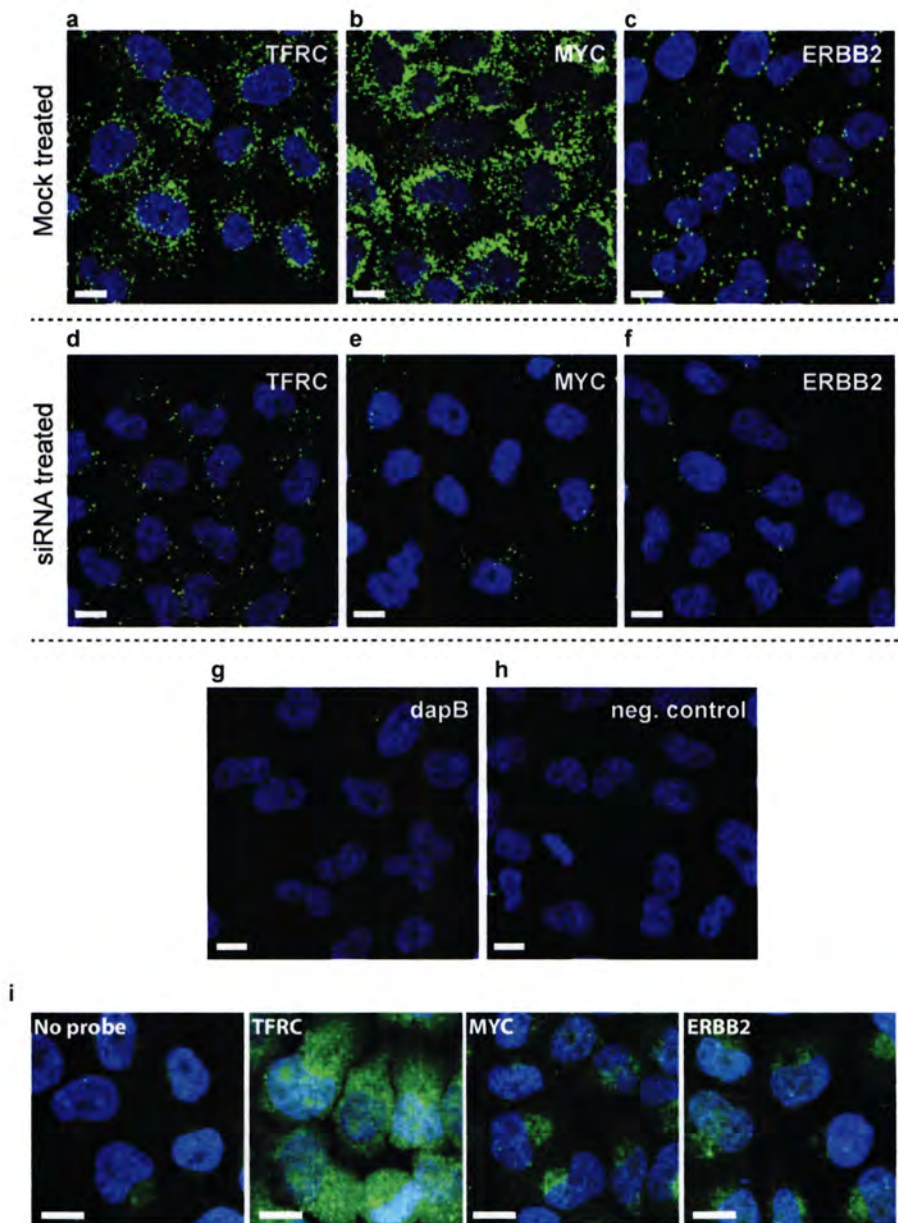
VA10-10829	BCR	613
VA10-10955	SLC17A7	57030
VA10-11037	MX1	4599
VA10-11247	CSPG4	1464

References

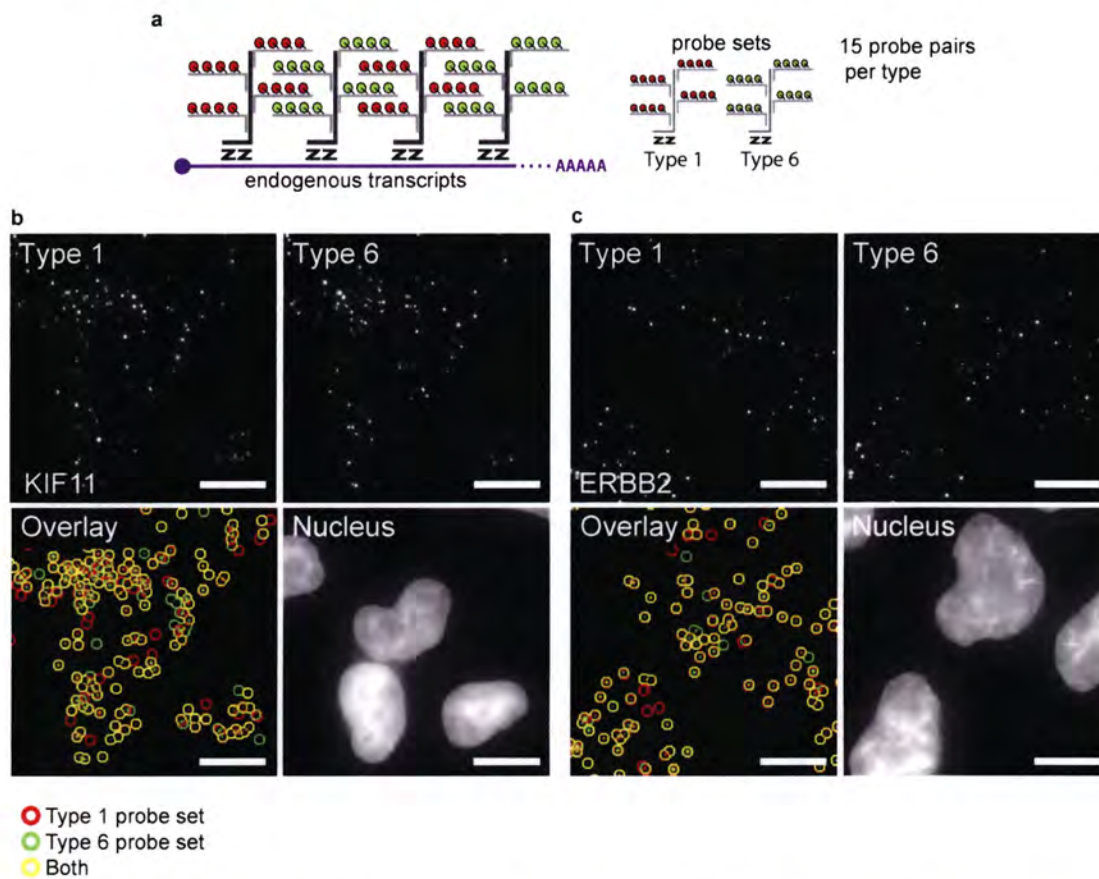
1. Powell, M. Direct search algorithms for optimization calculations. *Acta Numerica* (1998).
2. Elowitz, M., Levine, A., Siggia, E. & Swain, P. Stochastic gene expression in a single cell. *Science (New York, N.Y.)* **297**, 1183-6 (2002).
3. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods* **5**, 877-9 (2008).
4. Bertin, E. & Arnouts, S. SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement* (1996).
5. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* **7**, R100 (2006).
6. Martinez, W. L. & Martinez, A. R. *Computational Statistics Handbook with MATLAB, Third Edition.* (Chapman & Hall/CRC, London, UK, 2008)



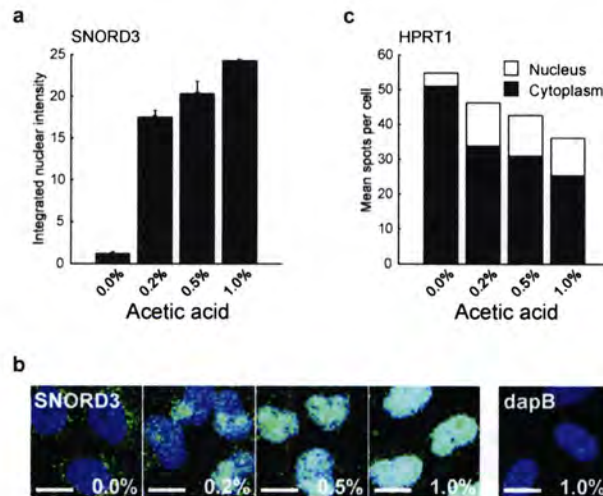
Supplementary Figure 1 | At similar imaging settings bDNA sm-FISH gives brighter signal than o-nuc sm-FISH. **(a)** Endogenous *MYC* transcript detected using bDNA sm-FISH (green) acquired with 100x magnification oil-immersion objective with 1ms exposure time and the camera gain set to one. The area marked in the lower panel corresponds to **Fig. 1e**. DAPI (cell nucleus) is blue. Scale bar: 13 μ m for main image and 5 μ m for insert. **(b)** As in **a**, but for o-nuc sm-FISH. Note that **a** and **b** were rescaled with the same parameters. **(c)** As in **b** but with a higher rescaling of the image. The area marked in the lower panel corresponds to **Fig. 1f**.



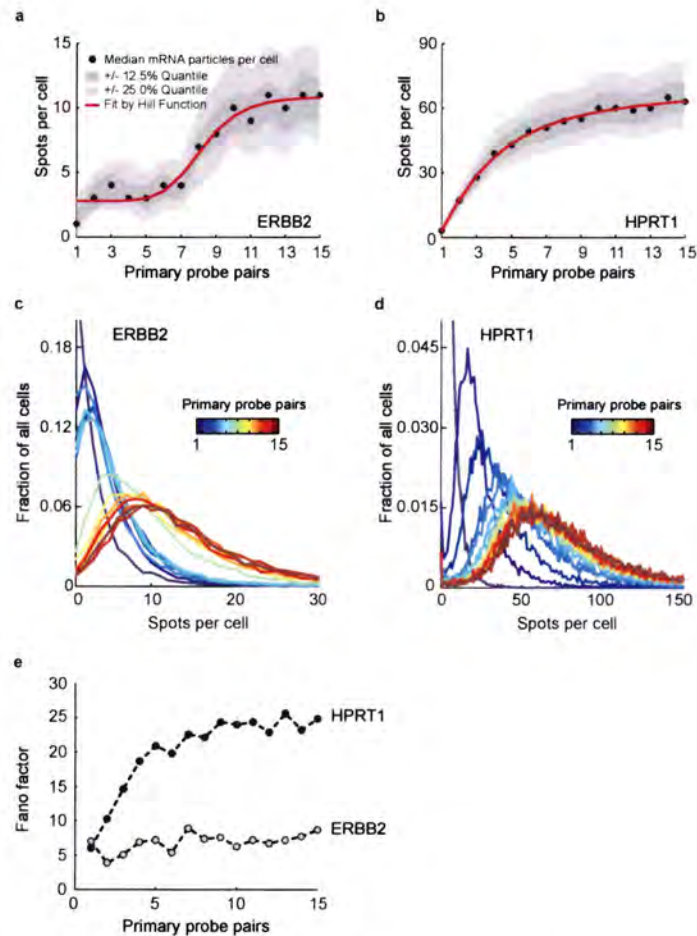
Supplementary Figure 2 | Signal from bDNA sm-FISH can be visualized with a high-throughput automated microscope and is specific for the targeted gene. HeLa cells were incubated for three days with siRNA targeting *TFRC*, *MYC* and *ERBB2*, then stained by bDNA sm-FISH (green) with probes targeting the respective genes. **(a-c)** Mock treated control cells hybridized with *TFRC*, *MYC* and *ERBB2* probes, respectively. **(d-f)** siRNA treated cells hybridized with *TFRC*, *MYC* and *ERBB2* probes, respectively. **(g,h)** Negative controls in mock treated cells using probes against the *E. coli* gene *dapB*, **g**, and a control without primary probe sets, **h**. **(i)** Acquisition of cells stained with o-nuc sm-FISH with probes against indicated transcripts and imaged with the high-throughput automated microscope (using 5 times longer exposure time than for bDNA sm-FISH). All images were acquired at 40x magnification with an automated spinning-disk microscope. DAPI (cell nucleus) is blue. Scale bars: 13μm.



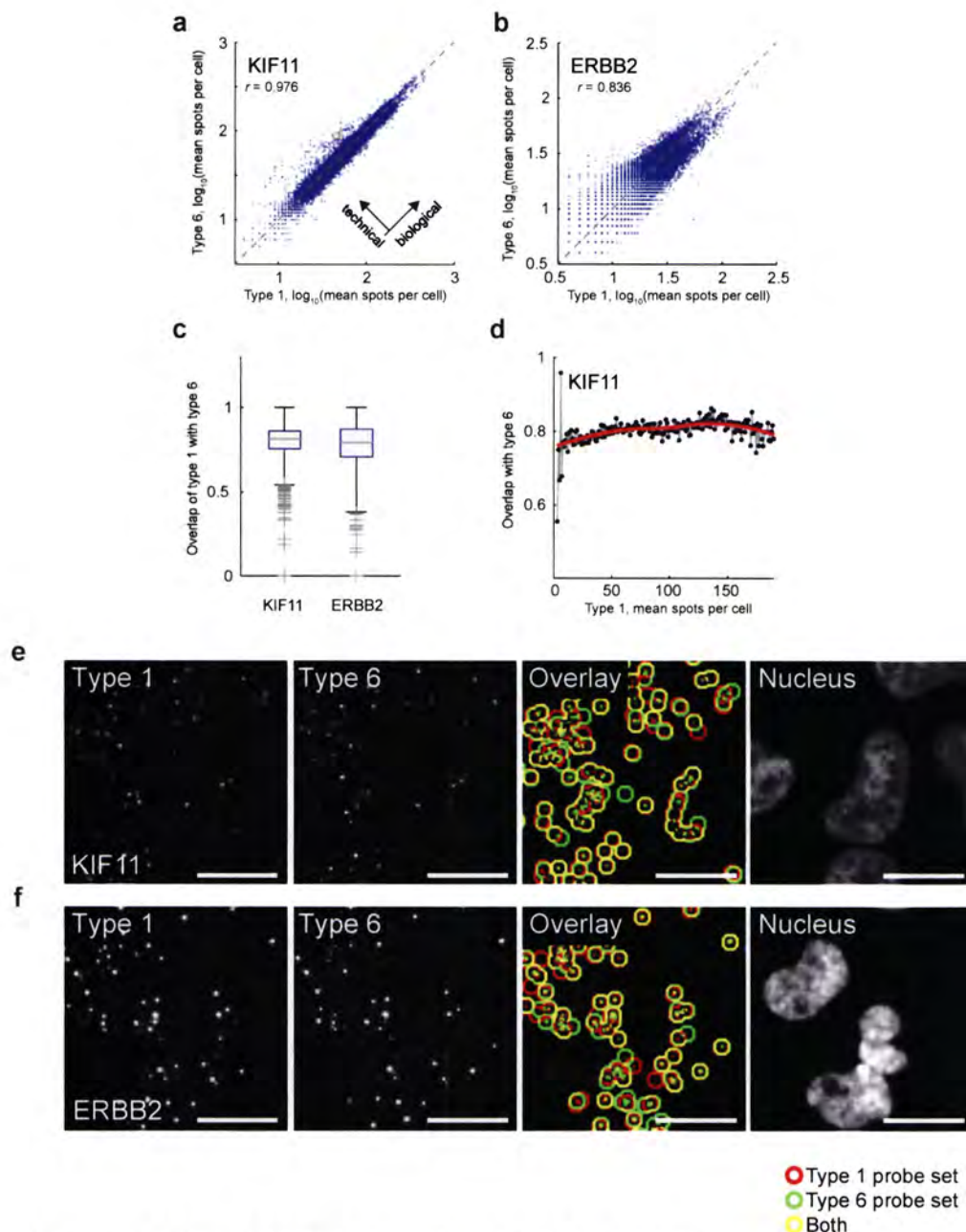
Supplementary Figure 3 | The overlap obtained by two bDNA FISH probe sets on endogenous transcripts is high. **(a)** Experimental setup with intercalating probe types along the transcript. Type 1 and 6 from Affymetrix were used. **(b,c)** Example images for the overlap obtained when *KIF11* and *ERBB2* endogenous transcripts were targeted using approach in **a**. Red circles: type 1 probe sets. Green circles: type 6 probe sets. Scale bar: 13 μ m.



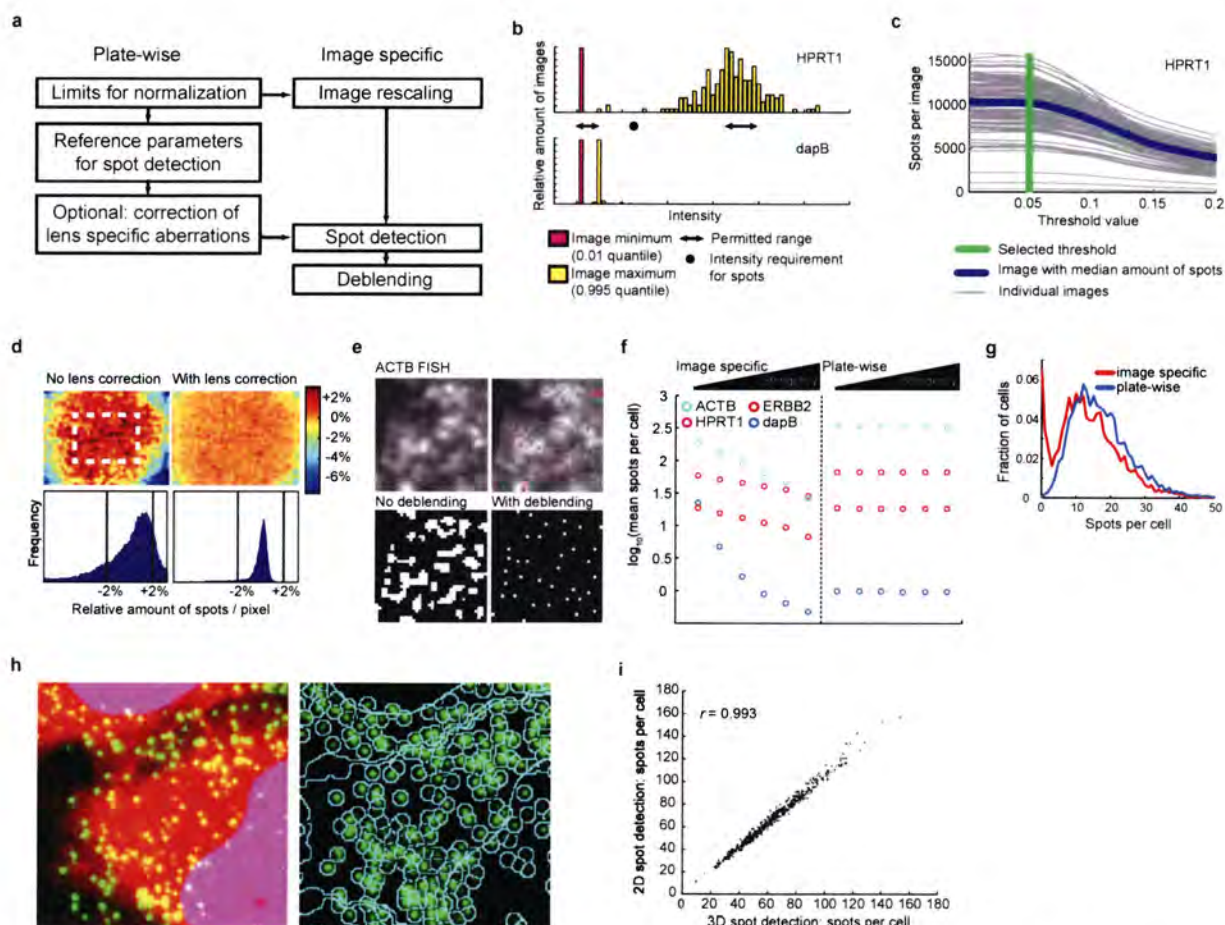
Supplementary Figure 4 | Acetic acid treatment improves nuclear accessibility but decreases staining in the cytoplasm. **(a,b)** Staining integrated intensity in the nucleus for *SNORD3* increases upon addition of acetic acid to the fixation solution (4% paraformaldehyde in PBS). Quantification of nuclear intensities is shown in **a**, while their corresponding example images are shown in **b**. bDNA sm-FISH is green, DAPI (cell nucleus) is blue. Error bars: s.d. Scale bar: 13 μ m. **(c)** Decrease of *HPRT1* cytoplasmic spots caused by addition of acetic acid to the fixation solution (back bars) and increase of *HPRT1* nuclear spots caused by addition of acetic acid to the fixation solution (white bars). Bar graph represents the mean of two replicates.



Supplementary Figure 5 | 10-15 probe pairs are required for accurate measurement of the mRNA number per cell using the bDNA FISH method. **(a)** The number of probe pairs targeting *ERBB2* was systematically increased (horizontal axis) and the number of mRNA spots per cell measured. Spot count per cell reached 80% saturation at nine probe pairs and 99% saturation at 15 probe pairs. **(b)** As in **a** but for the higher expressed transcript *HPRT1*. Spot count per cell reached 80% saturation at 10 probe pairs and 90% saturation at 15 probe pairs. **(c)** Distribution of mRNA spot counts per cell in samples stained with different numbers of probe pairs for *ERBB2*. **(d)** As in **c** but for *HPRT1*. **(e)** Fano factor (variance/mean) obtained for distributions shown in **c** and **d**.

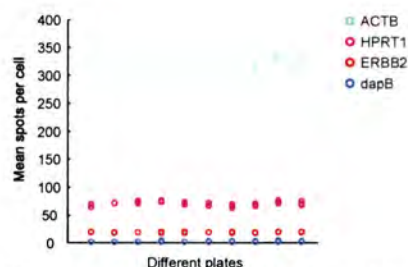


Supplementary Figure 6 | High-throughput bDNA sm-FISH is reproducible at the single cell level. **(a,b)** Correlation at the single cell level obtained targeting the same endogenous transcript with two different probe sets (**Supplementary Fig. 3a**) for *KIF11* and *ERBB2*, respectively. Pearson correlation value is shown, $n = 10,223$ cells and $n = 10,524$ cells for *KIF11* and *ERBB2*, respectively. The technical noise is given by the mean relative difference of spot counts per cell measured by type 1 and type 6 probe sets in a single cell, while the biological variability represents the difference of spot counts among different cells. **(c)** The fraction of detected spots by the type 1 probe set that were also detected by the type 6 probe sets in single cells, for *KIF11* and *ERBB2* transcripts, $n = 5,289$ and $n = 5,391$ cells, respectively. **(d)** The fraction of detected spots by the type 1 probe set that were also detected by the type 6 probe sets in single cells as a function of the type 1 spots per cells for *KIF11*. **(e,f)** Example images for the overlap obtained when *KIF11* and *ERBB2* endogenous transcripts were targeted using approach in **Supplementary Fig. 3a**. Red circles: type 1 probe sets. Green circles: type 6 probe sets. Scale bar: $13\mu\text{m}$. 37

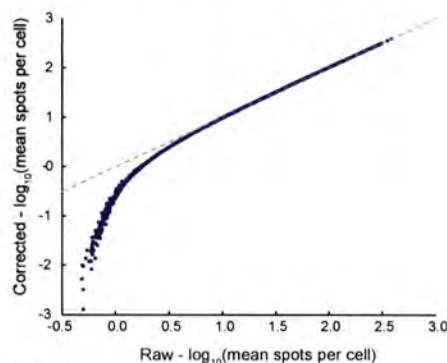


Supplementary Figure 7 | Plate-wise spot detection leads to a larger signal window between negative and positive controls and to more reliable transcript abundance distributions. **(a)** Representation of the plate-wise spot detection methodology. Limits for image rescaling, as well as reference parameters for spot detection and correction images for lens-specific aberrations are learned for each multi-well plate and applied to every image. **(b)** Distribution of image intensity minima and maxima for *HPRT1*, upper panel, and *E. coli dapB*, lower panel, are shown. Arrows indicate the permitted range for the image rescaling parameters, and the dot indicates the minimum intensity required for a pixel to be defined as a spot. **(c)** Selection of the threshold value for an example plate. The spot per image at different thresholds for a selection of images with cells hybridized with *HPRT1* probes are shown. **(d)** The bias introduced by lens aberrations is shown in the left image and histogram, the right image and histogram shows the after applying a filter to correct for position bias in the image. **(e)** Separation of spots detected in the high-expressed transcript *ACTB* using the SourceExtractor deblending algorithm. **(f)** Comparison of image specific and plate-wise spot detection. At similar levels of background spots in the *dapB* and no probes controls, plate-wise spot detection detects more spots for the three genes tested, *ACTB*, *HPRT1* and *ERBB2*, and is more robust to the stringency in the threshold. Image specific spot detection was carried out with the CellProfiler module *IdentifyPrimLog.m*. **(g)** The distribution of spots per cell was measured for *ERBB2* using image specific spot detections and plate-wise spot detections. The image-specific approach overestimates the number of cells with no mRNA spots. **(h)** Example of spot detection results in a single cell. bDNA sm-FISH is green, DAPI (cell nucleus) is blue, succinimidyl ester (cell outline) in red. **(i)** Comparison of 3D spot

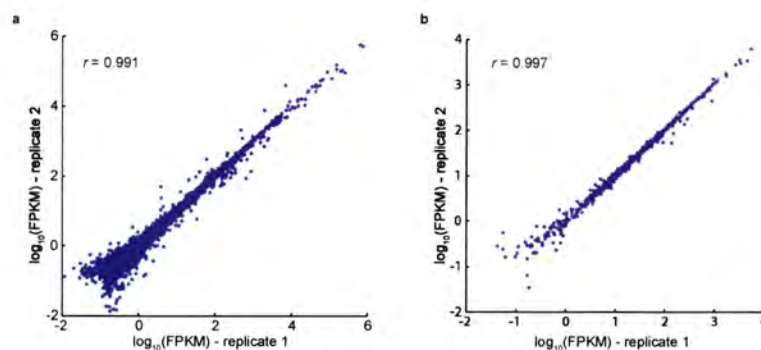
detection, i.e using 10 z-stacks, and 2D spot detection, using a maximum intensity projection image of the stacks. The Pearson correlation is 0.993 ($n = 15,238$), measurements from 500 randomly selected cells are shown.



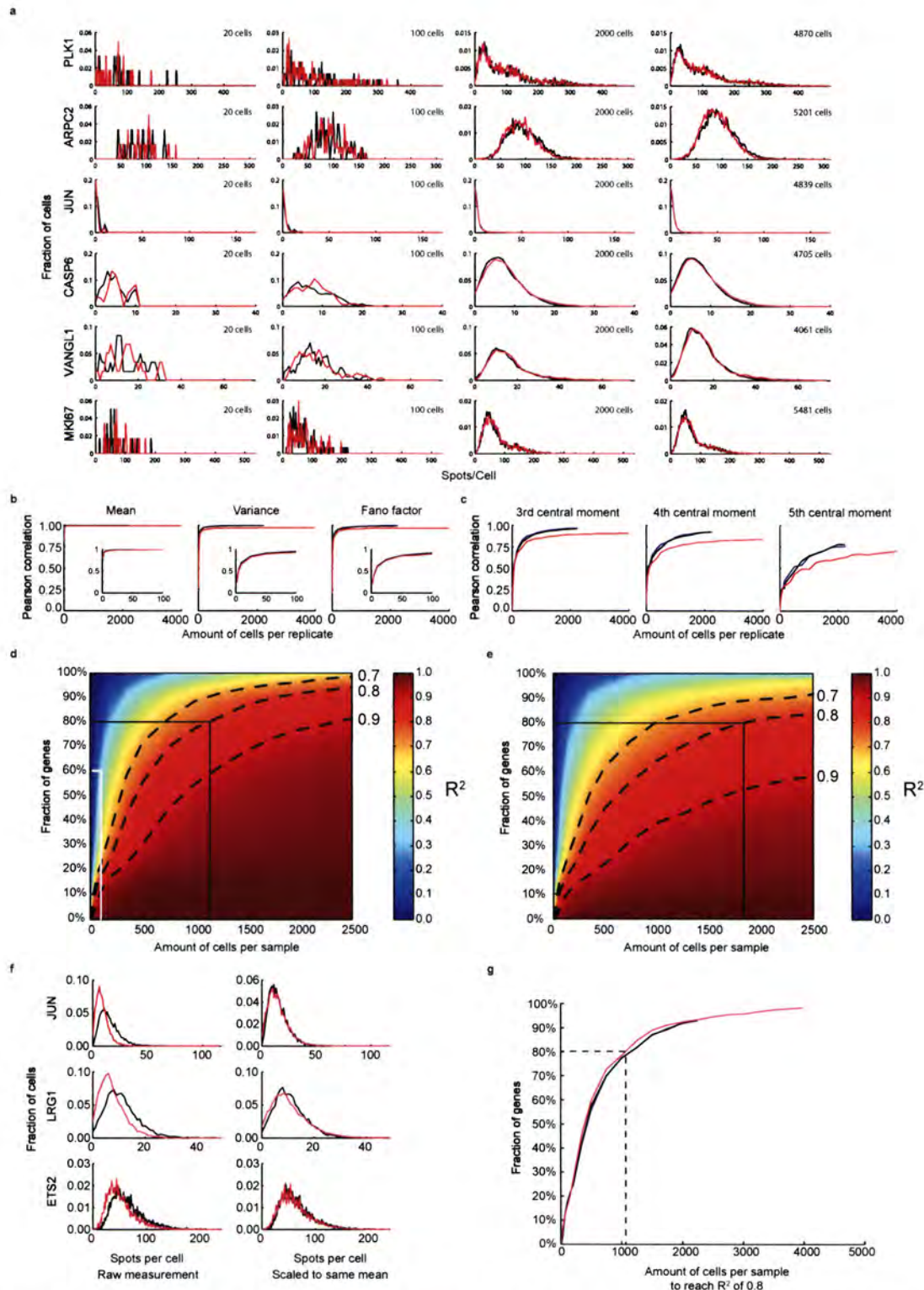
Supplementary Figure 8 | Gene expression levels of positive and negative controls are highly reproducible across plates in high-throughput bDNA sm-FISH. The logarithm of mean spots per cell in the positive controls (*ACTB*, *HPRT1* and *ERBB2*) and the negative control (*dapB*) is shown for different plates in the screen.



Supplementary Figure 9 | Background correction of mean spot count per cell only affects low-expressed genes. Correcting the mean spot count per cell by computing the fraction of cells above the *dapB* control is shown for the 928 genes targeted by the bDNA sm-FISH library. Broken line represent the isocline where the corrected $\log_{10}$ (mean spot per cell) equals the raw $\log_{10}$ (mean spots per cell) values.

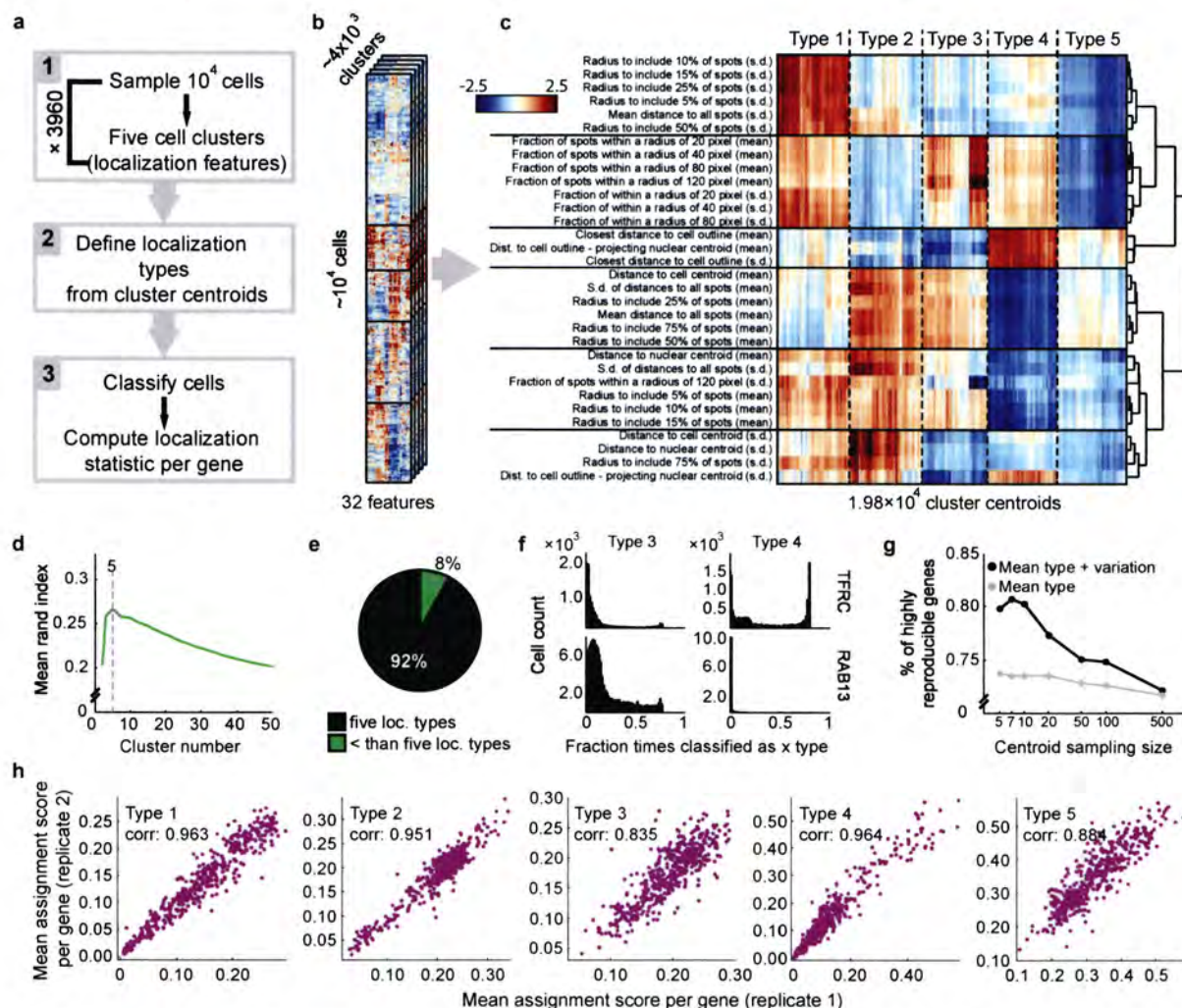


Supplementary Figure 10 | RNA-seq of HeLa cells is highly reproducible. **(a)** Reproducibility of RNA-seq values over the full data set. **(b)** Reproducibility of RNA-seq values over the 928 genes selected for the library. Pearson correlation values are shown.

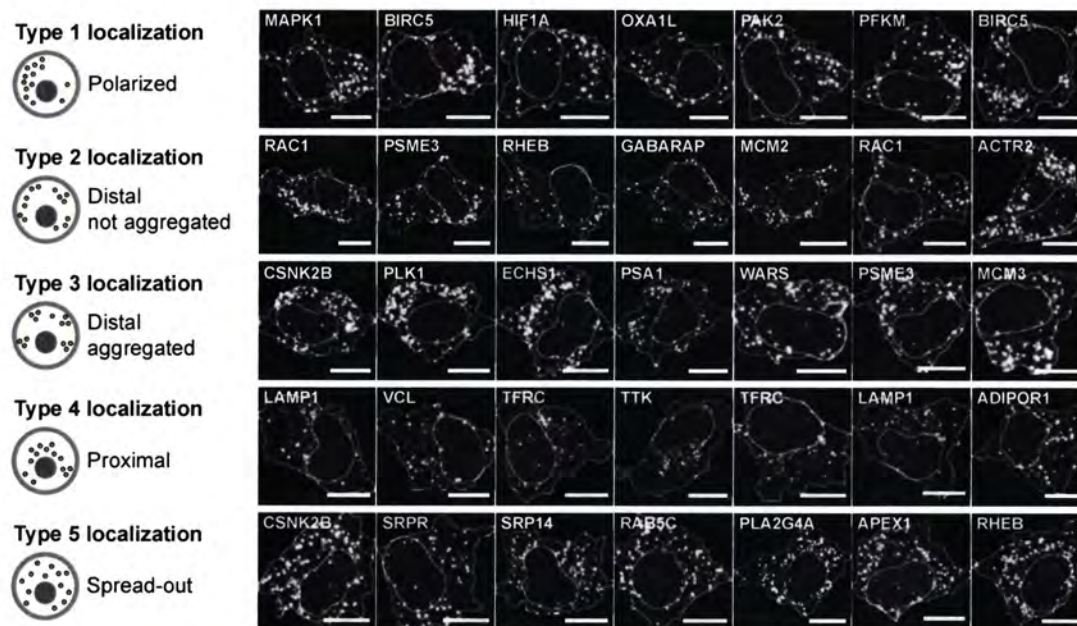


Supplementary Figure 11 | High-throughput bDNA sm-FISH reveals the minimum number of cells required for reproducible single-cell distributions of transcript abundance. **(a)** Example distributions of transcript abundance for two biological replicates (black and red lines) of different genes when sampling 20, 100, 2,000 or over 4,000 cells. **(b)** Reproducibility of the mean, variance and Fano factor (variance/mean)

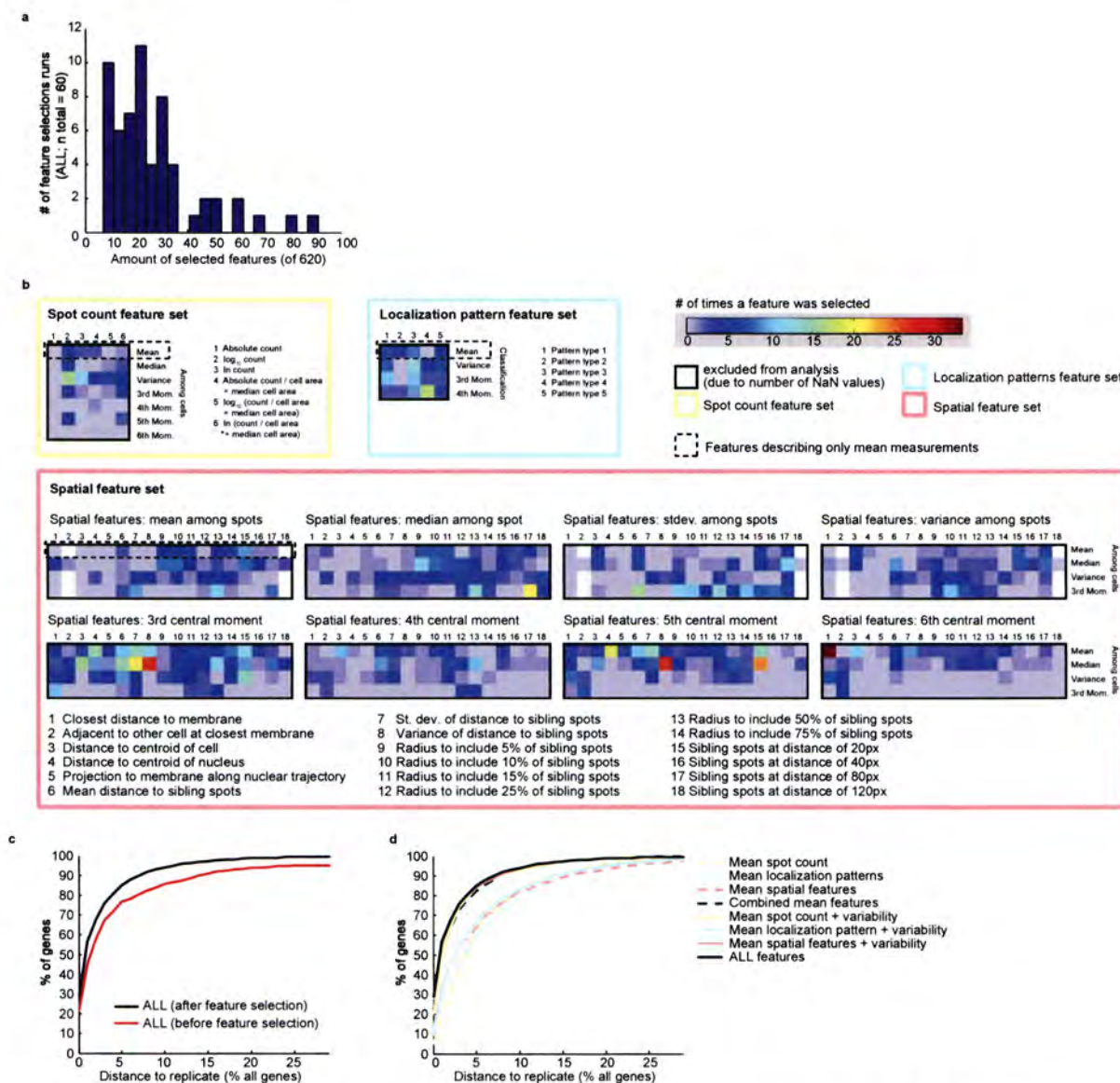
as function of the number of cell sampled comparing samples derived from the same replicates (blue and black lines) or two different replicates (red line). Lines represent the median of 100 bootstrapped samplings for each condition. **(c)** As in **b** but for the 3rd, 4th, and 5th, central moments. **(d-e)**. Coefficients of determination (R^2) between single-cell spot count distributions (at a bin size of 1 spot) comparing two distributions sampled within the same biological replicate, **d**, or between the two biological replicates, **e**. The percentage of genes showing a particular R^2 as a function of the number of cells sampled is shown. Isoclines for R^2 of 0.7, 0.8 and 0.9 are shown as broken lines. Black solid lines show the amount of cells required to achieve an R^2 of 0.8 or higher for ~80% of genes. White solid line in **d** shows the R^2 achieved (<0.5) for 60% of genes if only 100 are sampled. **(f)** The distribution shape between the two replicates is conserved after correction of spot per cells measurements with the mean spots number per cell of each replicate. **(g)** The isocline corresponding to an R^2 of 0.8 is shown for samplings within the same replicate (black and blue lines) and or between the two replicates after correction for differences in the mean spot per cell (red line). Lines represent the median of 100 bootstrapped samplings for each condition.



Supplementary Figure 12 | Identification of patterns of mRNA subcellular localization. (a) Approach taken to identify patterns of subcellular localization (**Supplementary Note 5**). (b) Example of hierarchical clustering result of $\sim 10^4$ sampled cells into five clusters, the sampling was repeated $\sim 4 \times 10^3$ times with replacement. This resulted in 1.98×10^4 cluster centroids when each hierarchical clustering was divided in five clusters, see d. (c) Definition of localization types by clustering of centroids from b. (d) Mean adjusted rand index at different partitions of clusters in b for two classifications of the same sampled cells ($n = 3,960$). Dashed line indicates the partition number at which the maximum adjusted rand index was achieved. (e) Percentage of clustering runs in b that show cluster centroids classified as five different pattern types (black) or less than five (green). (f) Distributions obtained at the cell population level for the fraction of times a single cell is classified as type 2 and type 4 for two example genes, *TFRC* and *RAB13*. (g) The effect of the centroid sampling size when assigning cell to a given localization type on the % of genes for which their transcript patterns are classified as reproducible (where the replicate gene is within the 10% closest genes of the replicate experiment). Taking into account the variability features describing the variability of the classification distributions in f (variance, 3rd and 4th central moments) increases reproducibility between replicates dramatically. Dashed lines are visual guides. (h) Correlation plots between the mean classification score for every pattern type between replicate 1 and 2.

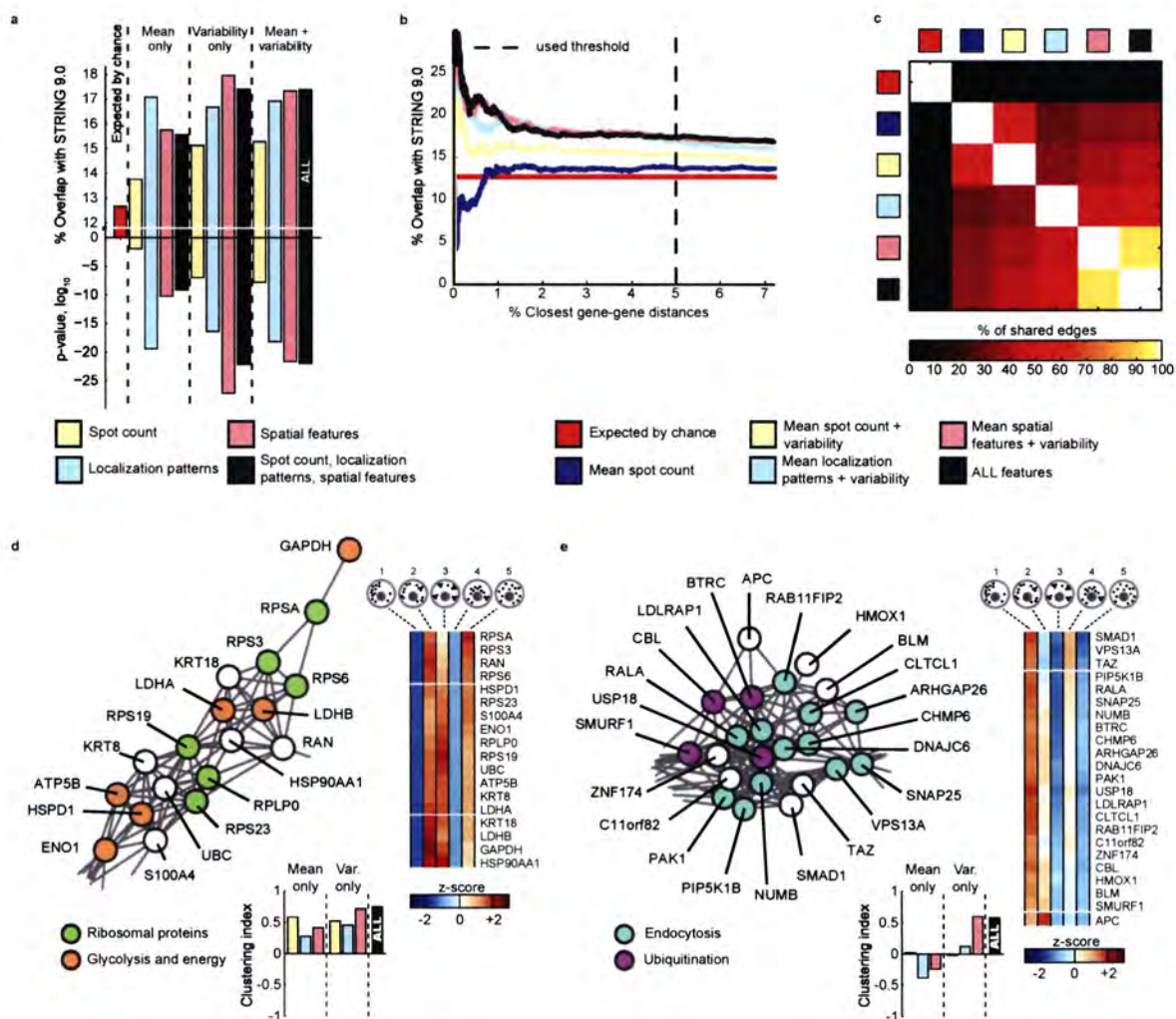


Supplementary Figure 13 | Interpretation of types of mRNA subcellular localization patterns. Left panel shows the interpretation of the five main types of subcellular localization patterns of transcripts determined in **Supplementary Fig. 11c**. The right panel shows seven examples of cells for every pattern type. All depicted cells were classified with the respective pattern type for the majority of classification iterations, and are within the 20 closest cells to the respective centroids of the pattern types in the 32-feature space used for clustering. bDNA sm-FISH is grey. Lines indicate cell segmentation. Scale bar: 13µm.



Supplementary Figure 14 | Feature selection based upon reproducibility. **(a)** Amount of features selected by individual iterations of feature selection using the feature set consisting of mean and variability features of spot count, localization patterns and spatial features (ALL features, see Fig. 5b). **(b)** Times a specific feature was selected shown for the ALL features starting set (n total = 60 selection rounds). Color boxes indicate the starting feature sets for analysis in **Fig. 5** and **Supplementary Fig. 14**: spot count (yellow), localization patterns (light blue), and spatial features (light red). Features highlighted with black dashed boxes show features describing only mean measurements. Central moments are abbreviated with Mom. **(c-d)** Fraction of genes for which the replicate pattern is within the indicated ranked distance to all genes of the replicate assay. Only genes within the testing sets of the feature selection procedure are considered (Online Methods). In **c** the reproducibility of genes using the ALL feature set before (red line) and after (black line) feature selection is shown. In **d** as in **c** but for distinct starting sets after feature selection, colors refer to **b**. Dashed lines consider only mean values of features (over both single cells and cell populations, features within black dashed boxes

in **b**. Solid lines consider mean and features of the distributions of primary spot feature values (both within single cells and across single cells) (variability features). Spot count variability refers to median, variance, and 3rd to 6th central moment between cells. Localization patterns variability refers to the variance, and 3rd and 4th central moment of the classification distributions. Spatial features variability refers to the median, standard deviation, variance, and 3rd to 6th central moment between spots and to the median, variance and 3rd central moment between cells.



Supplementary Figure 15 | Unbiased multivariate analysis of quantitative signatures of the *in situ* transcriptome in human tissue culture cells. **(a)** The overlap of the 5% smallest pairwise gene-gene distances [Euclidian distance between two genes based on their pairwise similarities (normalized Euclidian distance in feature space) with all other genes] with known gene interactions in STRING 9.0 and their respective p-values, calculated from various different sets of extracted features and 60 rounds of feature elimination to maximize reproducibility (Online Methods). Colors represent the type of feature sets used: 1) features of the spot count per cell (yellow); 2) features describing the types of mRNA localization patterns (light blue); 3) spatial features of spots (light red) and, 4) every feature type (black). The expected overlap by chance is also shown (red). In the first section of the bar graph only the mean values of features (over both single cells and cell populations) are considered. In the second section only features of the distributions of primary spot feature values (both within single cells and across single cells) (variability features) are considered. Spot count variability refers to median, variance, and 3rd to 6th central moment between cells. Localization patterns variability refers to the variance, and 3rd and 4th central moment of the classification distributions (Online Methods). Spatial features variability refers to the median, standard deviation, variance, and 3rd to 6th central moment between spots and to the median, variance and 3rd central moment between cells. In the third section, mean and variability features are

considered. ALL (the most right black bar) indicates the set used to derive the network in **Fig. 5b**. **(b)** The overlap of pairwise gene-gene distances with known gene interactions in STRING 9.0 is shown for the 0-7% top ranking smallest distances. Colors represent the used feature sets (see legend). Dashed line indicates 5%, the threshold used for connecting two genes and further analysis in **a**, **c-e**, and **Fig. 5**. **(c)** Percentage of shared gene-gene distances below the 5% threshold (network edges) between various feature sets (see legend). **(d,e)** Sub-region 3 and 4 of the network in **Fig. 5b**, showing a tight cluster of genes encoding for ribosomal proteins (green) and proteins involved in glycolysis/energy metabolism (orange), and a tight cluster of genes encoding for proteins involved in endocytosis (turquoise) or ubiquitination (purple). Also depicted are z-scored mean classification distributions of cells (as clustered heatmaps) for all five main types of single-cell spot localization patterns (Online Methods) for the genes in each sub-region. Type 1 corresponds to a polarized, type 2 to a distal, type 3 to a distal and aggregated, type 4 to a perinuclear, and type 5 to a spread-out spot distribution. Bar graphs indicate the clustering index for each feature type in each sub-region (color coding as in **b** and **c**).