HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2018 May 13.

Published in final edited form as:

Nat Methods. 2017 December ; 14(12): 1153–1155. doi:10.1038/nmeth.4500.

Profiling the transcriptome by RNA SPOTs

Chee-Huat Linus Eng[1], **Sheel Shah**[2], **Julian Thomassie**[3], and **Long Cai**[4]

[1]Division of Chemistry and Chemical Engineering, Caltech, Pasadena USA 91125

[2]UCLA-Caltech Medical Scientist Training Program, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, USA

[3]Division of Engineering and Applied Science, Caltech, Pasadena USA 91125

[4]Division of Biology and Biological Engineering, Caltech, Pasadena USA 91125

Abstract

Single molecule FISH (smFISH) has been the gold standard in quantifying individual transcripts abundances. Here, we demonstrate the scaling up of smFISH to the transcriptome level by profiling of 10,212 different mRNAs from mouse fibroblast and embryonic stem cells. This methods, called RNA SPOTs (Sequential Probing of Targets), provides an accurate and low-cost alternative to sequencing in profiling transcriptomes.

RNA sequencing (RNAseq)[1,2] has been a powerful method to quantify RNAs in a diverse range of biological samples. While RNAseq has replaced microarrays as the de-rigueur method for genomics studies because of higher sensitivities and dynamic range, reverse transcription and other steps needed to convert RNA to cDNA to sequencing libraries can introduce biases in the quantitation of mRNAs. Moreover, sequencing the RNAs at nucleotide level is not necessary for counting the abundances of transcripts. Single molecule fluorescence in situ hybridization (smFISH)[3,4], which directly hybridize DNA oligonucleotide probes to transcripts in cells, is highly sensitive and accurate in quantitating mRNA abundances.

Here, we demonstrate transcriptome level profiling of mRNAs with single molecule sensitivity and high accuracy using a method based on sequential FISH (seqFISH)[5]. We had shown that seqFISH can be applied to image hundreds of transcripts in cells and tissues[6], image dynamics of chromosomes[7] and allow lineage tracking with single cell resolution[8]. However, the major limitation of seqFISH is that optical diffraction limit prevents many mRNAs from being resolved simultaneously in single cells. In principle, super-resolution microscopy[9] and expansion microscopy[10] can resolve the optical density issue *in situ*.

Correspondence should be addressed to lcai@caltech.edu.

Contributions: C.-H.L.E. and L.C. conceived and designed experiments. C.-H.L.E. performed all the experiments. S.S. wrote the scripts for image analysis. C.-H.L.E. performed image analysis with the guidance of S.S. J.T. wrote the script for probe generations. C.-H.L.E. and L.C. performed data analysis. All authors contributed in writing the manuscripts.

Competing Financial Interests: L.C., C.-H.L.E., and S.S have filed a patent on RNA SPOTs and its barcoding scheme.

However, many applications quantify mRNAs that have been extracted from cells and tissues. In these cases, capturing transcripts onto an oligonucleotide dT surface and adjusting the dilution factors can easily remove the optical crowding problems and allow the transcriptome to be decoded by seqFISH.

To distinguish this *in vitro* application from the *in situ* seqFISH experiments, we call this approach RNA SPOTs (Sequential Probing Of Targets). Extracted mRNAs were first captured on a Locked Nucleic Acid(LNA) poly(dT) functionalized coverslip (Fig 1a) and then hybridized with a pool of 323,156 primary probes targeting the coding regions of 10,212 mRNAs with 28 to 32 probes each gene (Figure 1a-b, Supplementary Table 1 and **Online Methods)**. To barcode the 10,212 genes with sequential hybridization, we used a 12 "pseudo-color" based scheme such that 4 rounds of barcoding are sufficient to cover the transcriptome (12^4=20,736) (Supplementary Table 2), with an additional round of error correction to compensate for one drop in any round of barcoding[6] (Fig 1c-d). The pseudo-colors design shortens the number of barcoding rounds, which reduces the errors in reading out barcodes.

To implement the pseudo-color scheme, we designed the primary probes to contain a 25-nt RNA binding sequence, as well as 4 overhang sites[11] that can be bound by dye-labeled readout oligos (Figure 1b). Each site has 12 possible sequences corresponding to the 12 pseudo-colors. To readout the 12 pseudo-colors, three of the readout oligos were hybridized at a time, imaged in the Cy3b, Alexa 594 and Alexa 647 fluorescence channels, and repeated 4 times to iterate through all 12 readout sequences, with disulfide cleavage[12,13] in between the hybridizations to remove the fluorophores (Supplementary Fig 1 and 2).

With 5 rounds of barcoding using the 12 pseudo-color readouts scheme, a total of 60 readout oligos were used to decode the 10,212 genes targeted (Supplementary Fig 1-4 and Supplementary Table 3). Each set of primary probes that target a specific gene contain 5 unique readout sequences that are spread out over the overhang sites (Fig 1b). A total of 20 rounds of hybridization, or 5 barcoding round each containing 4 serial hybridizations (Supplementary Fig 1) were performed. A common sequence is present in all primary probes and targeted by an oligo labeled with Alexa 488 to serve as an alignment marker through all 20 rounds of hybridization (Supplementary Fig 2b). Each four rounds of serial hybridization were collapsed onto a single image with 12 pseudo colors (Fig 1c). The barcodes were determined from aligning five barcoding rounds of the pseudo-color images. The switching and rehybridization time is fast, with the overall speed limited by imaging speed. Typically, 100-200 fields of view containing more than 10^6 mRNAs can be imaged with 20 rounds of serial hybridization in a 14-hour period through an automated fluidics system. We use Spots per Millions (SPM) to normalize spots counts for individual genes between experiments (Supplementary Table 4 and 5). The false positive rates of detection is low, with 0.72 ± 1.9 SPM per barcode, as determined by the remaining 238,620 off-target barcodes.

To determine the accuracy of the transcriptome level measurements, we compare the decoded RNA SPOTs data with RNAseq data in mouse fibroblasts (NIH/3T3) and mouse embryonic stem cells (mESCs), and found that they correlated with R=0.86 and R=0.9

respectively (Fig 2a, b and Supplementary Fig 5 and Supplementary Table 4). Between two replicates of RNA SPOTs in fibroblasts, the results agree with R=0.94, indicating that RNA SPOTs is a highly robust and reproducible measurement method (Fig 2c, Supplementary Fig 5- 7). Finally, RNA SPOTs correlated with the gold standard smFISH quantitation with a correlation of R=0.86 in mESCs (24 genes)[14] and R= 0.88 in fibroblasts (7 genes) (Fig 2d and Supplementary Fig 8).

Comparing genes that were differentially expressed in fibroblasts versus mESCs, we observed the same trend as those detected by RNAseq. For example, pluripotency factors such as *Rex1* (also known as *Zfp42*), *Esrrb* and *Sox2* are highly expressed in mESCs but not expressed in fibroblasts as determined by RNA SPOTs. Similarly, genes involved in extracellular matrix maintenance, such as *Timp2*, *Timp3* and Collagen related genes such as *Col4a1*, *Col6a3* are up-regulated in fibroblast cells compared to mESCs (Fig 2e and Supplementary Table 6).

Another advantage of RNA SPOTs compared to RNAseq is that specific sets of genes can be profiled selectively. In this fashion, ribosomal RNA and highly expressed housekeeping genes can be avoided simply by eliminating those probes from the gene set. As each dot detected in our assay corresponds to a single mRNA, RNA SPOTs is more efficient in term of imaging compared to RNAseq, where many sequencing reads are needed to determine the abundance of a transcript. The current barcoding space is sufficient for the entire transcriptome, and noncoding RNAs and other RNAs without polyA tails can be captured in hydrogels (Supplementary Fig 9) rather than with dT oligos.

SPOTs is a significant improvement over existing Nanostrings technology[15] because of the genome level coverage and the higher specificity due to the larger number of probes used per gene. By incorporating amplification methods such as HCR [6,16], SPOTs signal can potentially allow faster imaging with air objectives and higher throughput comparable to RNAseq.

RNA SPOTs can be scaled down to single cell in combination with microfluidics tools to trap and lyse cells[17] or with split-pool molecular indexing methods[18]. While SPOTs cannot be used to discover new RNA sequences, identification of new cell types only require quantifying the combinatorial expression patterns of genes. Thus, there is no need to re-sequence the mRNAs at the nucleotide level just to count their abundances. With targeted RNA SPOTs, we can choose to probe only for the 2000 transcription factors[19] or 1000 landmark informative genes[20] in single cells, instead of profiling the transcriptome, to capture the essential information in cells and to increase the number of cells sampled. As cost of sequencing is a major limiting factor in many genomics experiments, SPOTs enable an accurate and low-cost alternative to sequencing with many further applications beyond RNA to DNA and proteins.

Online Methods

Primary probe design

Gene specific primary probes were designed as previously described with some modifications[6]. Probe sets were crafted separately for each gene and then refined as a full set to mitigate cross-hybridization in the experiment. Individual probe sets were first crafted using exons only from within the CDS region of the gene. For genes that did not yield enough targeted probes from the CDS region only, exons from both the CDS and 5′ UTR regions were used. The masked genome and annotation database from UCSC were used to look up the gene sequences. Consensus regions of all spliced isoforms were identified. 25-nt sequences of the gene sequences were extracted from these exons, and their GC contents were calculated. Probe sequences that fell outside of the allowed GC range (45-70% in this case) were immediately dropped. In addition, we dropped any probe sequences which contained 5 or more consecutive nucleotide bases of the same kind. A local BLAST query was run on each remaining probe against a BLAST database that was constructed from GENCODE reversed introns and mRNA sequences. BLAST hits on any sequences other than the target gene with a 15-nt match were considered off-target hits. We compiled a collection of RNA-seq data from ENCODE and computed a copy number table for all the genes across different samples. This off target copy number table was used to evaluate the off target hits. Any probe that hit an expected total off-target copy number exceeding 10,000 FPKM was dropped. Probes were sequentially dropped from genes until any off-target gene was hit by no more than 6 probes from entire pool. At this stage, all of the viable probes for the gene had been identified. For the final probe set, the best possible subset from the viable probes was selected such that none of the final probes were within 2 nucleotide bases of each other on the target sequence. The overlapping probes were grouped and sorted by distance from the target GC content (55% in this case). Overlapping probes were removed in order of descending distance from target GC, starting from the probe with the greatest distance, until no overlaps remained. To minimize cross hybridization between probe sets, a local BLAST database was constructed from all the viable probe sequences, and the probes were queried against it. All matches of 17-nt or longer between probes were removed by dropping the matched probe from the larger probe set. For this experiment, the targeted probe set size range was set to 28-32 probes. Any probe set with more than 32 probes was trimmed down by removing probes with the farthest GC content from 55%. To design the 20-nt readout sequences, a set of probe sequences were randomly generated with the 4 bases nucleotides. Readout probe sequences with range 45-60% GC were selected. We used BLAST to eliminate any sequences that matched with any contiguous homology sequences longer than 14-nt to the mouse transcriptome. The reverse complements of these readout sequences were included in the primary probes according to the designed barcodes.

Primary probe construction

Primary probes were ordered as an oligoarray complex pools from Twist Bioscience and were constructed as previously described [6,11,21]. Briefly, a 2-step limited PCR cycles were used to amplify the designated probe sequences from the oligo complex tool. Then, the amplified products were purified using QlAquick PCR Purification Kit (28104; Qiagen) according to the manufacturer's instructions. The PCR products were used as the template

for in vitro transcription (E2040S; NEB) followed by reverse transcription (EP7051; Thermo Fischer) with the forward primer. After alkaline hydrolysis, the single stranded DNA (ssDNA) probes were purified by ethanol precipitation and resuspend in primary probe hybridization buffer comprising of 30% formamide (F9037; Sigma), 2× SSC (15557036; Thermo Fischer), and 10% (w/v) Dextran Sulfate (D8906; Sigma). The probes were stored at -20°C.

Readout probe synthesis

20-nt readout probes were ordered from Integrated DNA Technologies (IDT) as 3′ thiol modified at its oxidized form. Alexa Fluor 647 Cadaverine (A30679; Invitrogen) and Alexa Fluor 594 Cadaverine (A30678; Invitrogen) were reacted with *N*-Succinimidyl 3-(2-pyridyldithio)propionate, SPDP (P3415; Sigma) at 1: 100 ratio in 1× PBS (AM9624, Ambion) at room temperature for at least 4 hours on a shaker. Then, the mixture was purified using PD MiniTrap G-10 (28-9180-10; GE Healthcare), and was evaporated in a vacuum concentrator. The dye-linker intermediate product was kept at -20°C until the conjugation with 3′ thiol oligonucleotide probes. 10mM TCEP (77720; Thermo Scientific) was used to activate the 3′ thiol readouts at 37°C for 30 minutes. Then the oligonucleotides were purified using illustra NAP-5 columns (17-0853-02; GE Healthcare), and the oligonucleotides were directly eluted in 1× PBS with 10mM EDTA (15575020; Thermo Fischer) and were mixed with the dye-linker intermediate product. The reaction was allowed to proceed at room temperature for 2 hours. Then, the mixture was ethanol precipitated, HPLC purified, resuspend into 500nM concentration in 1× Tris-EDTA buffer (93283; Sigma) and was kept at -20°C. To conjugate Cy3B fluorophore (PA63101; GE Healthcare) to the 3′ thiol oligonucleotides, a (3-(2-pyridyldithio)propionyl hydrazide), PDPH (22301; Thermo Scientific) was used instead of the SPDP linker.

Coverslips functionalization

Coverslips were functionalized as previously described[17] with some modifications. Briefly, coverslips (3421; Thermo Scientific) were sonicated in 100% ethanol for 20 minutes. After drying, the coverslips were cleaned with a plasma cleaner at HIGH (PDC-001, Harrick Plasma) for 5 minutes. Then, the coverslips were immediately immersed in a 2% (v/v) trimethoxysilane aldehyde (PSX1050; UCT Specialties) solution made in pH 3.5 10% (v/v) acidic ethanol solution for 15 minutes at room temperature. After triple rinsing of the coverslips with ethanol, the coverslips were heat-cured at 90°C for 10 minutes. Then an oligonucleotide reaction mixture containing 2.5 μM 5′-aminated LNA-oligo(dT) (300100-02; Exiqon), cyanoborohydride coupling buffer (C4187; Sigma), and 1M sodium chloride (AM9759; Thermo Fischer) was sandwiched between two coverslips at room temperature in a humid hybridization chamber for 3 hours. The coverslips were then rinsed with Millipore water and dried with compressed air. A quenching reaction mixture made from 10%(v/v) 100mM pH7.5 Tris-HCl (15567027; Thermo Fischer) buffer in cyanoborohydride coupling buffer was added to the entire silanized surface of the coverslips to quench the remaining aldehyde functional groups at room temperature for 30 minutes. Finally, the coverslips were rinsed with water and dried with compressed air. All coverslips were made fresh before SPOTs experiment.

Cell cultures and RNA Preparation

Mouse ES-E14 cells were cultured as previously described[14]. Mouse NIH/3T3 cells (ATCC) were cultured in DMEM (10569044; Gibco) supplemented with 10% FBS (S11150; Atlanta biologicals) and 1% penicillin (10378016; Gibco). Once the cell confluency reached 60-80%, the total RNA was extracted using RNeasy Mini Kit (74104; Qiagen) according to the manufacturer's instructions.

Hydrogel immobilization

Coverslips were first sonicated at 100% ethanol for 20 minutes, followed by plasma cleaning with a plasma cleaner at HIGH for 5 minutes. The coverslips were then immersed in the 2% PlusOne bind-silane(17-1330-01) solution made in ethanol for 30 minutes at room temperature. After rinsing the coverslips with ethanol for several times, the coverslips were dried at 90°C for 30 minutes. Purified total RNA was mixed in 4% acrylamide/bis solution (1610147; Bio-Rad) with fresh 25mM VA-044 initiator (27776-21-2; Wako Chemical) and the solution was degassed for 10 minutes on ice. A 12mm square coverslip (470019-000; VWR) was functionalized with GelSlick (Lonza; 50640). 1uL of the RNA hydrogel solution was added to the bind-silane functionalized coverslip and was spread out using the GelSlick functionalized square coverslip. The thickness of the hydrogel formed can be controlled by manipulating the volume added. The polymerization happened in a humid hybridization at 37°C for 2 hours. After polymerization is complete, the coverslips were immersed in 2× SSC for an hour or more to facilitate the removal of the top coverslips. smFISH measurement was then performed according to standard protocol.

Primary probe hybridization

A custom Secure Seal Flowcell, 2 × 28mm 3mm ID, 35 × 15 OD, 0.25mm thick (RD478685-M; Grace Bio-labs) was applied on the functionalized poly(dT) coverslips. For NIH/3T3 cells experiments, 50 ng of total RNA in RNA binding buffer comprising of 1M LiCl (L9650; Sigma), 40mM pH7.5 Tris-HCl, 2mM EDTA, 0.1% Triton X-100 (93443; Sigma), and 20U of SUPERase IN RNase Inhibitor (AM2694, Ambion) was allowed to be captured at room temperature for 1 hour. For ES-E14 experiment 1 and 2, the amount of total RNA used was 50 ng and 5 ng respectively. Once the mRNA is immobilized on the coverslip, 20 uL of 1 nM/probe for a total of 323,156 probes in hybridization buffer containing 30% formamide (F9037; Sigma), 2× SSC (15557036; Thermo Fischer), and 10% (w/v) Dextran Sulfate (D8906; Sigma) was hybridized to the targeted mRNA at 37°C for 24 hours in a humid hybridization chamber. After hybridization, the sample was washed for 30 minutes at room temperature with wash buffer containing 40% formamide, 2× SSC, and 0.1% Triton X-100 to remove non-specific binding of the primary probes. The sample preparation of primary probe hybridization ended with a 3 times washes with 2× SSC and was kept in 2× SSC until the next step.

RNA SPOTs imaging

Each readout probes hybridization mixture contained 10nM each for three unique readout probes either conjugated to Alexa 647, Alexa 594 or Cy3b in hybridization buffer comprising 10% formamide, 2× SSC, and 10% (w/v) Dextran Sulfate (D4911; Sigma). Each

serial hybridization takes 15 minutes to achieve optimal fluorescent signals, followed by a 4-minutes high stringency wash containing 20% formamide and 2× SSC to remove non-specific binding of probes. Once the first hybridization is complete, the flow cell was connected to an automated fluidics delivery system made from two multichannel fluidics valves (EZ1213-820-4; IDEX Health & Science) and a peristaltic pump (NE-9000G-UP, New Era Pump Systems Inc.). The integration of the fluidics valves, peristaltic pump, and microscope imaging were controlled through a custom script written in Micromanager software. Once the flow cell is connected, ~100 to ~200 frame of views (FOVs) were imaged at 647-nm, 594-nm, 532-nm, and 488-nm channels with 500 ms exposure time under anti-bleaching buffer containing 20mM Tris-HCl pH 8 (15568025; Thermo Fischer), 50mM NaCl, 3mM Trolox (238813; Sigma), 0.8% glucose (G7528; Sigma), 3U/mL pyranose oxidase (P4234; Sigma) or 50U/mL of glucose oxidase (G2133; Sigma), and 20 U/mL SUPERase IN RNase Inhibitor. The anti-bleaching buffer was stored under a layer of mineral oil (M5904; Sigma) throughout the whole experiment. Imaging was done using a standard epifluorescence microscope (Nikon Ti Eclipse with custom built laser assembly), a Nikon 60× oil objective and a sCMOS camera (Zyla 4.2; Andor). Nikon Ti Eclipse PFS autofocus was activated to keep the plane focused during imaging. Once the imaging is complete, reduction buffer made from 50mM TCEP (646547; Sigma), 2× SSC, and 0.1% Triton X-100 was flowed into the flow cells and the solution was allowed to incubate for 5 minutes. Then, 2× SSC buffer supplemented with 20U/mL SUPERase IN RNase Inhibitor was flowed into the flow cell in excess for 4minutes to completely remove the TCEP solutions. As our flow cell only takes ~22uL of solution, 200uL of subsequent serial hybridization solutions was flowed into the flow cell each time to ensure hybridization. The whole process was repeated until 20 rounds of hybridizations were imaged. Generally, a SPOTs experiment takes ~14 hours for imaging 100-200 FOVs. After the SPOTs imaging is complete, a few FOVs were imaged to use for threshold and illumination background corrections in image analysis. A multispectral beads slide was imaged at the end of experiment for chromatic aberration corrections.

Image Processing

To remove the effects of chromatic aberration, multispectral beads were first used to create geometric transforms to align all fluorescence channels. Next, the background illumination profile of every fluorescence channel was mapped using a morphological image opening with a large structuring element. These illumination profile maps were used to flatten the illumination in post-processing resulting in relatively uniform background intensity and preservation of the intensity profile of fluorescent points. The background signal was then subtracted using the ImageJ rolling ball background subtraction algorithm with a radius of 3 pixels. Finally, the calculated geometric transforms were applied to each channel respectively.

Image Registration

As the Alexa 488 channel labeled all the spots in the field of view, this channel was used to align all sets of images using a normalized 2D image cross-correlation.

Barcode calling

The potential RNA signals were then found by finding local maxima in the image above a predetermined pixel threshold in the registered images. Once all potential points in all channels of all hybridizations were obtained, dots were matched to potential barcode partners in all other channels of all other hybridizations using a 1-pixel search radius to find symmetric nearest neighbors. Point combinations that constructed only a single barcode were immediately matched to the on-target barcode set. For points that matched to construct multiple barcodes, first the point sets were filtered by calculating the residual spatial distance of each potential barcode point set and only the point sets giving the minimum residuals were used to match to a barcode. If multiple barcodes were still possible, the point was matched to its closest on-target barcode with a Hamming distance of 1. If multiple on target barcodes were still possible, then the point was dropped from the analysis as an ambiguous barcode. This procedure was repeated using each hybridization as a seed for barcode finding and only barcodes that were called similarly in at least 4 out of 5 rounds were used in the analysis. The number of each barcode was then counted and transcript numbers were assigned based on the number of on-target barcodes present. The remaining barcodes were used to assess the false positives rate by running through the same process. All image processing and image analysis code can be obtained upon request.

smFISH

Unless stated, all smFISH measurements were conducted with 1nM/probe concentration with a total number of 24 probes targeting a gene in hybridization buffer comprising of 10% formamide, 2× SSC and 10% (w/v) dextran sulfate at 37°C. The probes were conjugated to either Alexa 647, Alexa 594, or Cy3b dyes. NIH/3T3 cells were fixed with 4% paraformaldehyde (28908; Sigma) in 1x PBS at room temperature for 10 minutes. After washes with 1x PBS, the cells were permeabilized using 70% ethanol and kept in -20°C. The probe sequences for each gene were designed using Stellaris Biosearch Technologies and the probes were ordered from IDT with 5′ amine modifications. The probes were conjugated to dye as previously described [Lubeck 2014]. After hybridization, the sample was washed with wash buffer supplemented with 30% formamide and 2× SSC at room temperature for 30 minutes. The samples were then stained with DAPI (D1306; Thermo Fischer) in 2× SSC, followed by imaging under anti-bleaching buffer. The cells were segmented and the copy numbers for each gene were counted using a custom Matlab script.

RNA-Seq

RNA-seq data[22] were obtained from Gene Expression Omnibus (GEO) with an accession number of GSE98674. Briefly, the total RNA was purified using RNeasy Mini Kit following the manufacturer's instruction. The library was constructed using NEBNext ultra RNA-seq (E7530; NEB) according to the manufacturer's instructions and sequenced on Illumina HiSeq2500. Base calls were performed with RTA 1.13.48.0 followed by conversion to FASTQ with bcl2fastq 1.8.4. Alignment was performed using TopHat algorithm. Transcript assembly and FPKM estimates were done using Cufflinks algorithm.

Statistics and Reproducibility

The technical replicates for RNA SPOTs of NIH/3T3 and ES-E14 are 2 in both cell cultures. The R values in the plots of technical replicates and SPOTs versus RNAseq are Pearson's r correlation coefficient. For smFISH average measurements, the error bars represent the standard error of the mean. For differential gene expression analysis, two-tailed student t-test is carried out with $n = 2$ for mean SPM for both NIH/3T3 and ES-E14. *P-values* smaller than 0.05 and log2 fold change greater and less than ± 2 respectively are used as a threshold for significance.

A “Life Sciences Reporting Summary” for this publication is available.

Data and Custom Codes Availability

The raw data and custom scripts from this study are available from the corresponding author upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank P. Sims, A. Raj, and B. Wold for discussions. We thank Y. Zhang for help with automated fluidics system, N. Koulena for helping in probe generation, and E. Lubeck for help at early stage of the project. This work was supported by grants from US National Institutes of Health (NIHRNA-1-NIHP.HD075605 and RNASUPP-1-NIHP.HD075605).

References

1. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008 Jul; 5(7):621–8. [PubMed: 18516045]
2. Nagalakshmi U, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008; 320:1344–1349. [PubMed: 18451266]
3. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006; 4:e309. [PubMed: 17048983]
4. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of Single RNA Transcripts in Situ. Science. 1998; 280:585–590. [PubMed: 9554849]
5. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. Single-cell in situ RNA profiling by sequential hybridization. Nat Methods. 2014; 11:360–361. [PubMed: 24681720]
6. Shah S, Lubeck E, Zhou W, Cai L. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron. 2016; 92:342–357. [PubMed: 27764670]
7. Takei Y, Shah S, Harvey S, Qi LS, Cai L. Multiplexed Dynamic Imaging of Genomic Loci by Combined CRISPR Imaging and DNA Sequential FISH. Biophys J. 2017 May 9; 112(9):1773–1776. [PubMed: 28427715]
8. Frieda KL, Linton JM, Hormoz S, Choi J, Chow KK, Singer ZS, Budde MW, Elowitz MB, Cai L. Synthetic recording and in situ readout of lineage information in single cells. Nature. 2017 Jan 5; 541(7635):107–111. [PubMed: 27869821]
9. Lubeck E, Cai L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat Methods. 2012; 9:743–748. [PubMed: 22660740]
10. Chen F, Tillberg PW, Boyden ES. Expansion microscopy. Science. 2015; 347:543–548. [PubMed: 25592419]

11. Beliveau BJ, Joyce ER, Apostolopoulos N, Yilmaz F, Fonseka CY, McCole RB, Chang Y, Li JB, Senaratne TN, Williams BR, Rouillard JM, Wu Ct. A versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proc Nat Acad Sci USA. 2012; 109:21301–6. [PubMed: 23236188]
12. Daniel, Steven G., et al. FastTag nucleic acid labeling system: a versatile method for incorporating haptens, fluorochromes and affinity ligands into DNA. RNA and oligonucleotides Biotechniques. 1998; 24:484–489. [PubMed: 9526662]
13. Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc Natl Acad Sci U S A. 2016 Sep 27; 113(39):11046–51. [PubMed: 27625426]
14. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, Cai L, Elowitz MB. Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol Cell. 2014 Jul 17; 55(2):319–31. [PubMed: 25038413]
15. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T, Ratcliffe AL, Webster PJ, Davidson EH, Hood L, Dimitrov K. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008 Mar; 26(3):317–25. [PubMed: 18278033]
16. Choi HM, Beck VA, Pierce NA. Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. ACS Nano. 2014 May 27; 8(5):4284–94. [PubMed: 24712299]
17. Bose S, Wan Z, Carr A, Rizvi AH, Vieira G, Pe'er D, Sims PA. Scalable microfluidics for single-cell RNA printing and sequencing. Genome Biol. 2015 Jun 6.16:120. [PubMed: 26047807]
18. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, Adey A, Waterston RH, Trapnell C, Shendure J. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science. 2017 Aug 18; 357(6352):661–667. [PubMed: 28818938]
19. Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, Sladek R. TFCat: the curated catalog of mouse and human transcription factors. Genome Biol. 2009; 10(3):R29.doi: 10.1186/gb-2009-10-3-r29 [PubMed: 19284633]
20. Donner Y, Feng T, Benoist C, Koller D. Imputing gene expression from selectively reduced probe sets. Nat Methods. 2012 Nov; 9(11):1120–5. [PubMed: 23064520]
21. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M. The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X-chromosome. Science. 2013
22. Antebi YE, Linton JM, Klumpe H, Bintu B, Gong M, Su C, McCardell R, Elowitz MB. Combinatorial Signal Perception in the BMP Pathway. Cell. 170:1184–1196.e1124.

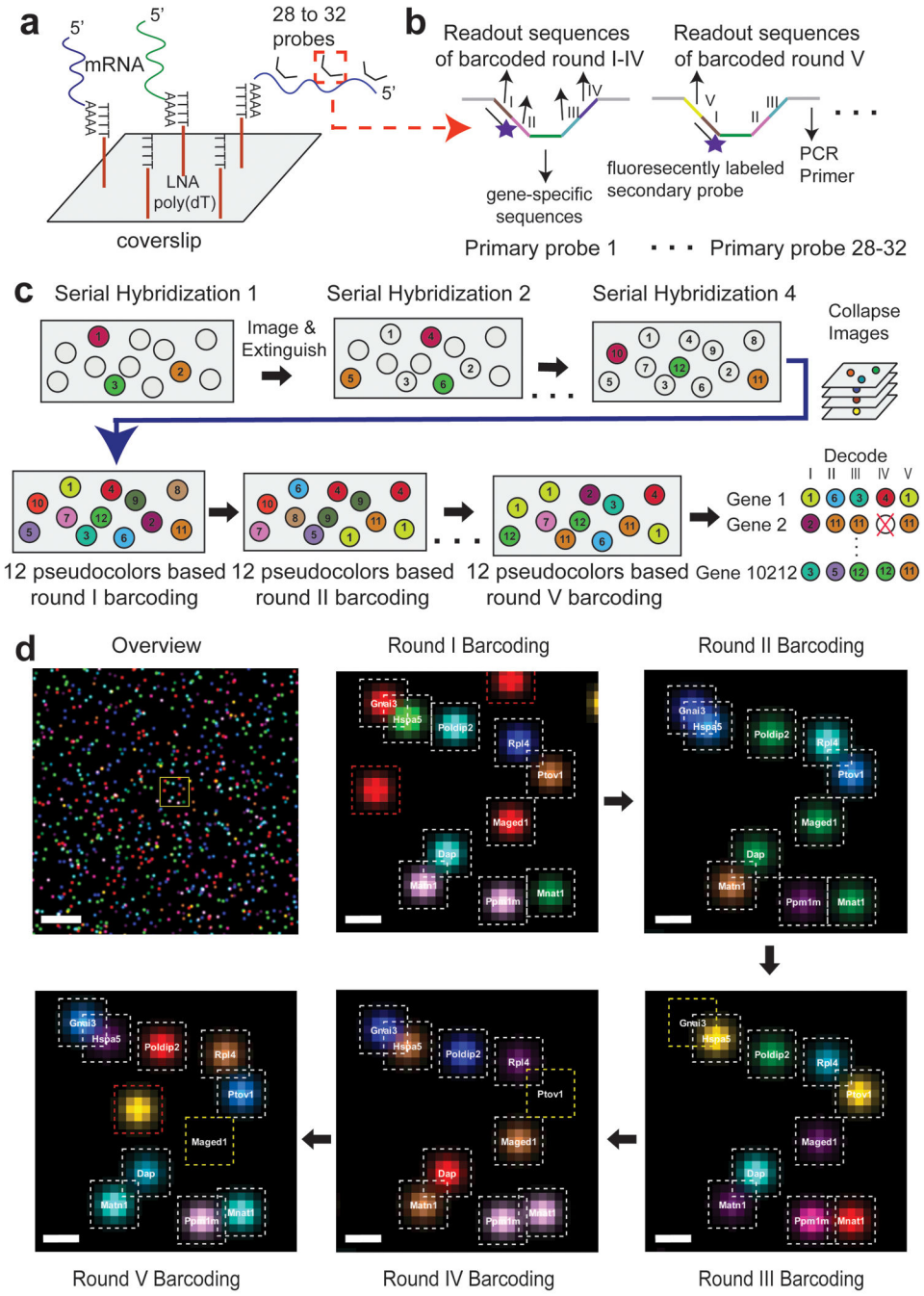

Figure 1.

RNA SPOTs profiles 10,212 mRNAs *in vitro*. (a) mRNA is first captured on a Locked Nucleic Acid(LNA) polyd(T) functionalized coverslip. Next, gene specific primary probes (323,156 total) are hybridized against the 10,212 targeted mRNA. Each gene is targeted by 28-32 primary probes. (b) Primary probes structure. Each 149-mer primary probe comprises a 25-nt gene specific sequences complementary to the mRNA, 4 out of 5 barcoded rounds sequences (20-nt each), which are read out by fluorescently labeled secondary readout probes, spacers of nucleotide 'T' added between readout sequences and gene specific

region(1-nt each), and 2 primers binding sites (20-nt each) (c) In each serial hybridization, 3 readout probes conjugated to either Alexa 647, Alexa 594, or Cy3b fluorophore are hybridized to the primary probes and imaged. Then the fluorescent signals are extinguished, followed by the next serial hybridization. Images from 4 rounds of serial hybridization are then collapsed into 1 composite image to generate a 12 “pseudocolor” image of round I barcoding. Images from the next 4 rounds of serial hybridization are collapsed to form the second composite image of round II barcoding and so on. 20 rounds of serial hybridizations generate a total of 5 composite images each containing 12 “pseudocolors”. The mRNA spots are decoded by aligning 5 composite images. The barcodes were generated from 12^4 = 20736 codes, with an extra round of error correction for mishybridization. (d) Digitized composite images based on actual experiments to decode 10,212 distinct mRNA. (White dashed squares: correctly identified barcodes. Red dashed squares: false positive; yellow dashed squares: barcodes identified despite mis-hybridization in one round of hybridization. Scale bars: Overview- 10μm; Round I to V barcoding- 1μm)

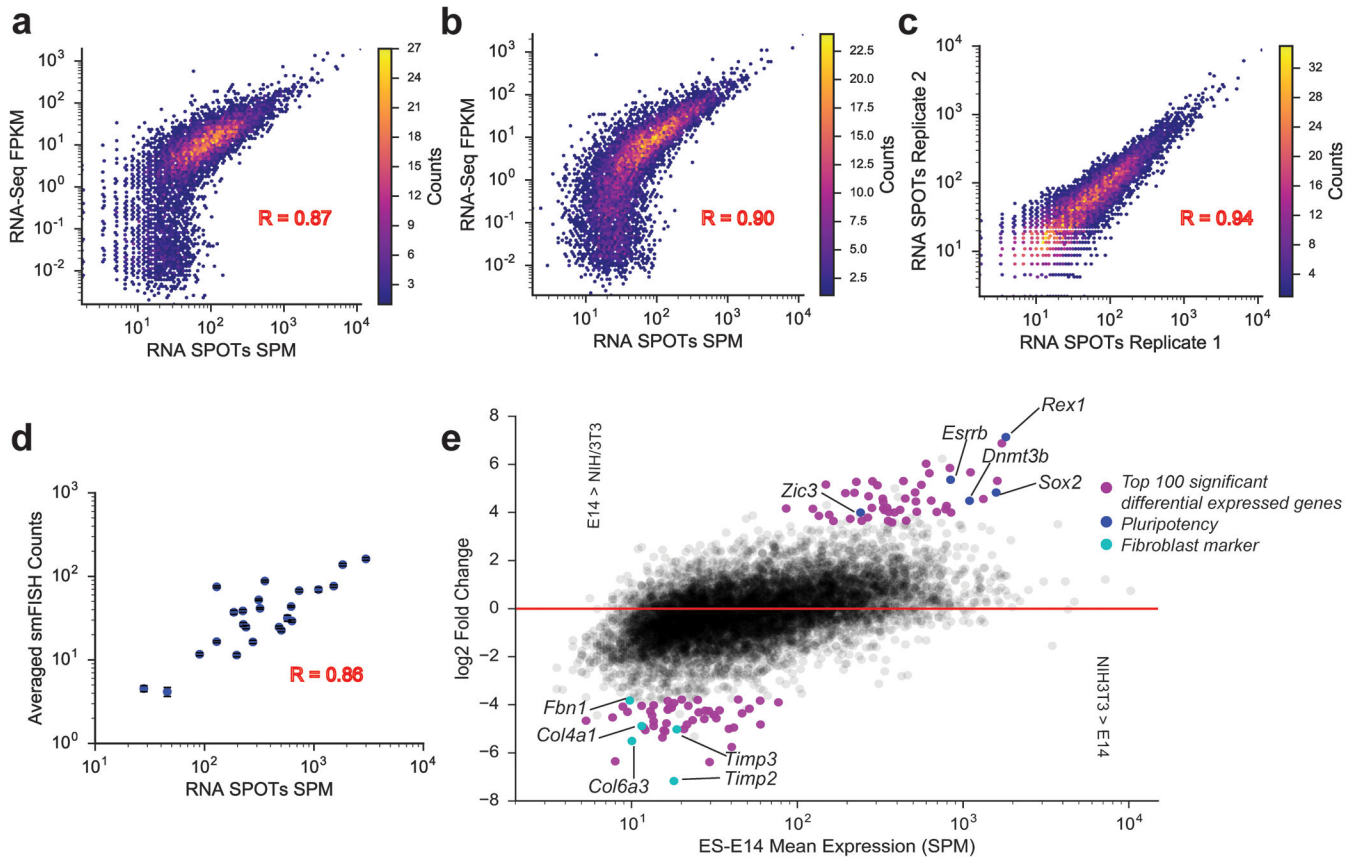

Figure 2.

RNA SPOTs is highly accurate and efficient. (a) Transcriptomic profiling of mouse NIH/3T3 cells by RNA SPOTs correlates strongly with measurement from RNA-seq. SPM (spots per million) normalizes the number of each decoded mRNA spots (n = 581,772) by the total number of spots. FPKM, fragments per kilobase per million reads (b) RNA SPOTs profiling of mouse ES-E14 cell line agrees strongly with RNA-seq measurement. (n = 1,688,747 spots) (c) Comparison of two RNA SPOT replicates profiling NIH/3T3 cells illustrates that the method is highly reproducible (n_1 = 581,772 spots; n_2 = 453,679 spots) (d) Comparison of averaged smFISH copy numbers of 24 genes in ES-E14 cells with RNA SPOTs SPM verifies the high accuracy measurement of SPOTs. Error bars represent standard errors of the mean (SEM) across different measurements in single cells. (e) Differential gene expression between NIH/3T3 and ES-E14 cells. *P-values* smaller than 0.05 as determined from two-tailed student t-test and log2 fold change greater and less than ± 2 respectively are used as a threshold for significance. Magenta dots represent top 50 upregulated and top 50 downregulated genes between the two cell lines. Blue dots represent the well-known genes involved in pluripotency. Cyan dots represent the genes involved in extracellular matrix maintenance.