

or apoptosis. As such, small-molecule inhibitors of RIPK3 may have limited therapeutic benefit because of their potential to promote apoptotic cell death.

Nec-1 blocked necroptosis induced by TNF/zVAD.fmk (fig. S5C), was protective in an inflammatory disease model (15), and does not induce apoptotic cell death, suggesting that inhibition of RIPK1 rather than RIPK3 may have therapeutic benefit. To mimic RIPK1 inhibition in the whole animal, we generated *Ripk1* knock-in mice with Asp¹³⁸ of the HKD motif necessary for kinase activity mutated to Asn (fig. S6, A and B). Whereas *Ripk1*^{+/+} mice died soon after birth, *Ripk1*^{D138N/D138N} mice were viable. Consistent with a critical role for the kinase activity of RIPK1 in TNF-induced necroptosis, *Ripk1*^{D138N/D138N} bone marrow–derived macrophages (BMDMs) and E1A-immortalized MEFs were as resistant as *Ripk3*^{+/+} cells to killing by TNF/zVAD.fmk (Fig. 4A and fig. S6C). *Ripk1*^{D138N/D138N} cells expressed normal amounts of RIPK3 and MLKL (Fig. S6D). *Ripk1*^{D138N/D138N} mice also resembled *Ripk3*^{+/+} mice in their systemic response to TNF, exhibiting less hypothermia than did their wild-type counterparts (Fig. 4B). Unlike RIPK1 loss, RIPK1 D138N did not rescue embryonic lethality caused by RIPK3 D161N (fig. S4). This result is consistent with nec-1 not protecting MEFs ex-

pressing RIPK3 D161N (fig. S5B) and indicates that the kinase activity of RIPK1 is not required for activation of caspase-8 by RIPK3 D161N.

RIPK1 is required for TNF-induced nuclear factor κ B (NF- κ B) and mitogen-activated protein kinase signaling (16, 17). Wild-type RIPK1 and RIPK1 D138N restored these signals in *Ripk1*^{+/+} cells (18), suggesting that RIPK1 has an essential scaffold function in this setting, whereas its kinase activity is dispensable. Indeed, *Ripk1*^{D138N/D138N} and wild-type BMDMs were indistinguishable in their phosphorylation of inhibitor of NF- κ B α (I κ B α), c-Jun N-terminal kinase (JNK), and extracellular signal-regulated kinase (ERK) in response to TNF (Fig. 4C). These results, together with the viability of *Ripk1*^{D138N/D138N} mice, are encouraging because they suggest that inhibiting the kinase activity of RIPK1 has no deleterious effects, at least in the short term. These *Ripk1*^{D138N/D138N} mice can be used to explore the contribution of RIPK1 and necroptosis to various mouse models of human disease.

References and Notes

1. K. Moriawaki, F. K. Chan, *Genes Dev.* **27**, 1640–1649 (2013).
2. Y. S. Cho et al., *Cell* **137**, 1112–1123 (2009).
3. S. He et al., *Cell* **137**, 1100–1111 (2009).
4. D. W. Zhang et al., *Science* **325**, 332–336 (2009).
5. N. Takahashi et al., *Cell Death Dis.* **3**, e437 (2012).
6. K. Newton, X. Sun, V. M. Dixit, *Mol. Cell Biol.* **24**, 1464–1469 (2004).
7. E. E. Varfolomeev et al., *Immunity* **9**, 267–276 (1998).
8. A. Oberst et al., *Nature* **471**, 363–367 (2011).
9. W. J. Kaiser et al., *Nature* **471**, 368–372 (2011).
10. L. Salmena et al., *Genes Dev.* **17**, 883–895 (2003).
11. X. Sun, J. Yin, M. A. Starovasnik, W. J. Fairbrother, V. M. Dixit, *J. Biol. Chem.* **277**, 9505–9511 (2002).
12. Y. H. Park, M. S. Jeong, H. H. Park, S. B. Jang, *Biochim. Biophys. Acta* **1834**, 292–300 (2013).
13. M. Muzio et al., *Cell* **85**, 817–827 (1996).
14. J. Seibler et al., *Nucleic Acids Res.* **31**, e12 (2003).
15. L. Duprez et al., *Immunity* **35**, 908–918 (2011).
16. M. A. Kelliher et al., *Immunity* **8**, 297–303 (1998).
17. T. H. Lee et al., *Mol. Cell Biol.* **23**, 8377–8385 (2003).
18. T. H. Lee, J. Shank, N. Cusson, M. A. Kelliher, *J. Biol. Chem.* **279**, 33185–33191 (2004).

Acknowledgments: We thank J. Starks, K. O'Rourke, A. Maltzman, M. Schlatter, T. Soukup, L. Tam, R. Pattini, R. Newman, H. Ngu, and D. Siler for technical assistance. S. Sun, W. Alexander, S. Akira, and Y. He kindly provided knockout mice. All authors were employees of Genentech.

Supplementary Materials

www.sciencemag.org/content/343/6177/1357/suppl/DC1
Materials and Methods
Figs. S1 to S6
References (19–25)

5 December 2013; accepted 6 February 2014
Published online 20 February 2014;
10.1126/science.1249361

Highly Multiplexed Subcellular RNA Sequencing in Situ

Je Hyuk Lee,^{1,2,*} Evan R. Daugherty,^{1,2,4,*} Jonathan Scheiman,^{1,2} Reza Kalhor,² Joyce L. Yang,² Thomas C. Ferrante,¹ Richard Terry,¹ Sauveur S. F. Jeanty,¹ Chao Li,¹ Ryoji Amamoto,³ Derek T. Peters,³ Brian M. Turczyk,¹ Adam H. Marblestone,^{1,2} Samuel A. Inverso,¹ Amy Bernard,⁵ Prashant Mali,² Xavier Rios,² John Aach,² George M. Church^{1,2,†}

Understanding the spatial organization of gene expression with single-nucleotide resolution requires localizing the sequences of expressed RNA transcripts within a cell in situ. Here, we describe fluorescent in situ RNA sequencing (FISSEQ), in which stably cross-linked complementary DNA (cDNA) amplicons are sequenced within a biological sample. Using 30-base reads from 8102 genes in situ, we examined RNA expression and localization in human primary fibroblasts with a simulated wound-healing assay. FISSEQ is compatible with tissue sections and whole-mount embryos and reduces the limitations of optical resolution and noisy signals on single-molecule detection. Our platform enables massively parallel detection of genetic elements, including gene transcripts and molecular barcodes, and can be used to investigate cellular phenotype, gene regulation, and environment in situ.

The spatial organization of gene expression can be observed within a single cell, tissue, and organism, but the existing RNA localization methods are limited to a handful of genes per specimen, making it costly and laborious to localize RNA transcriptome-wide (1–3). We originally proposed fluorescent in situ sequencing (FISSEQ) in 2003 and subsequently developed methods to sequence DNA amplicons on a solid substrate for genome and transcriptome sequencing (4–7); however, sequencing the cellular RNA in situ for gene expression profiling re-

quires a spatially structured sequencing library and an imaging method capable of resolving the amplicons.

We report here the next generation of FISSEQ. To generate cDNA amplicons within the cell (fig. S1), RNA was reverse-transcribed in fixed cells with tagged random hexamers (fig. S2A). We incorporated aminoallyl deoxyuridine 5'-triphosphate (dUTP) during reverse transcription (RT) (fig. S2B) and refixed the cells using BS(PEG)9, an amine-reactive linker with a 4-nm spacer. The cDNA fragments were then circularized before rolling

circle amplification (RCA) (fig. S2C), and BS(PEG)9 was used to cross-link the RCA amplicons containing aminoallyl dUTP (fig. S2, D and E). We found that random hexamer-primed RT was inefficient (fig. S3A), but cDNA circularization was complete within hours (fig. S3, B to D). The result was single-stranded DNA nanoballs 200 to 400 nm in diameter (fig. S4A), consisting of numerous tandem repeats of the cDNA sequence. BS(PEG)9 reduced nonspecific probe binding (fig. S4B), and amplicons were highly fluorescent after probe hybridization (fig. S4C). As a result, the amplicons could be rehybridized many times, with minimal changes in their signal-to-noise ratio or position (fig. S4, D and E). Using SOLiD sequencing by ligation (fig. S5), the signal overlap over 27 consecutive sequencing reactions was ~600 nm in diameter (fig. S4F). In induced pluripotent stem (iPS) cells, the amplicons counterstained subcellular structures, such as the plasma membrane, the nuclear membrane, the nucleolus, and the chromatin (Fig. 1A, fig. S6, and movies S1 to S3). We were able to generate RNA sequencing libraries in different cell types, tissue

¹Wyss Institute, Harvard Medical School, Boston, MA 02115, USA. ²Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. ³Department of Stem Cell and Regenerative Biology, Harvard University, Boston, MA 02138, USA. ⁴Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. ⁵Allen Institute for Brain Science, Seattle, WA 98103, USA.

*These authors contributed equally to this work.

†Corresponding author. E-mail: jehyuklee@mac.com (J.H.L.); gchurch@genetics.med.harvard.edu (G.M.C.)

sections, and whole-mount embryos for three-dimensional (3D) visualization that spanned multiple resolution scales (Fig. 1, B and C).

High numerical aperture and magnification are essential for imaging RNA molecules in single cells (8–10), but many gene expression patterns are most efficiently detected in a low-magnification and wide-field mode, where it typically becomes difficult to distinguish single molecules because of the optical diffraction limit and low sensitivity (11). To obtain a spot density that is high enough to yield statistically significant RNA localization, and yet sufficiently low for discerning individual molecules, we developed partition sequencing, in

which preextended sequencing primers are used to reduce the number of molecular sequencing reactions through random mismatches at the ligation site (Fig. 2A). Progressively longer sequencing primers result in exponential reduction of the observed density, and the sequencing primer can be changed during imaging to detect amplicon pools of different density.

Fluorescence microscopy can be accompanied by tissue-specific artifacts and autofluorescence, which impede accurate identification of objects. If objects are nucleic acids, however, discrete sequences, rather than the analog signal intensity, can be used to analyze the image. For FISSEQ,

putative nucleic acid sequences are determined for all pixels. The sequencing reads are then compared with reference sequences, and a null value is assigned to unaligned pixels. With a suitably long read length (L), a large number of unique sequences (n) can be used to identify transcripts or any other objects with a false-positive rate of approximately $n/4^L$ per pixel. Because the intensity threshold is not used, even faint objects are registered on the basis of their sequence, whereas background noise, autofluorescence, and debris are eliminated (Fig. 2B).

We applied these concepts to sequence the transcription start site of inducible mCherry mRNA in situ, analogous to 5' rapid amplification of cDNA ends–polymerase chain reaction (RACE-PCR) (12). After RT and molecular amplification of the 5' end followed by fluorescent probe hybridization (fig. S7A), we quantified the concentration- and time-dependent mCherry gene expression in situ (fig. S7B). Using sequencing-by-ligation, we then determined the identity of 15 contiguous bases from each amplicon in situ, corresponding to the transcription start site (fig. S7C). When the sequencing reads were mapped to the vector sequence, 7472 (98.7%) amplicons aligned to the positive strand of mCherry, and 3967 (52.4%) amplicons mapped within two bases of the predicted transcription start site (fig. S7D).

We then sequenced the transcriptome in human primary fibroblasts in situ (Fig. 3A) and generated sequencing reads of 27 bases with a median per-base error rate of 0.64% (fig. S8). Using an automated analysis pipeline (fig. S9), we identified 14,960 amplicons with size >5 pixels, representing 4171 genes, of which 13,558 (90.6%) amplicons mapped to the correct annotated strand (Fig. 3B, fig. S10, and table S1). We found that mRNA (43.6%) was relatively abundant even though random hexamers were used for RT (Fig. 3C). Ninety genes with the highest expression counts included fibroblast markers (13), such as fibronectin (*FNI*); collagens (*COL1A1*, *COL1A2*, *COL3A1*); matrix metalloproteinases and inhibitors (*MMP14*, *MMP2*, *TIMP1*); osteonectin (*SPARC*); stanniocalcin (*STC1*); and the bone morphogenesis–associated transforming growth factor (TGF)–induced protein (*TGFB1*), representing extracellular matrix, bone development, and skin development [Benjamini-Hochberg false discovery rate (FDR) $<10^{-19}$, 10^{-5} , and 10^{-3} , respectively] (Fig. 3D) (14). We made Illumina sequencing libraries to compare FISSEQ to RNA-seq. Pearson's r correlation coefficient between RNA-seq and FISSEQ ranged from 0.52 to 0.69 ($P < 10^{-16}$), excluding one outlier (*FNI*). For 854 genes with more than one observation, Pearson's r was 0.57 ($P < 10^{-16}$), 0.47 ($P < 10^{-16}$), and 0.23 ($P < 10^{-3}$) between FISSEQ and RNA-seq from fibroblasts, lymphocytes, and iPS cells, respectively (Fig. 3E). When FISSEQ was compared with gene expression arrays, Pearson's r was as high as 0.73 ($P < 10^{-16}$) among moderately expressed genes, whereas genes with low or high expression levels correlated poorly ($r < 0.4$) (fig. S11).

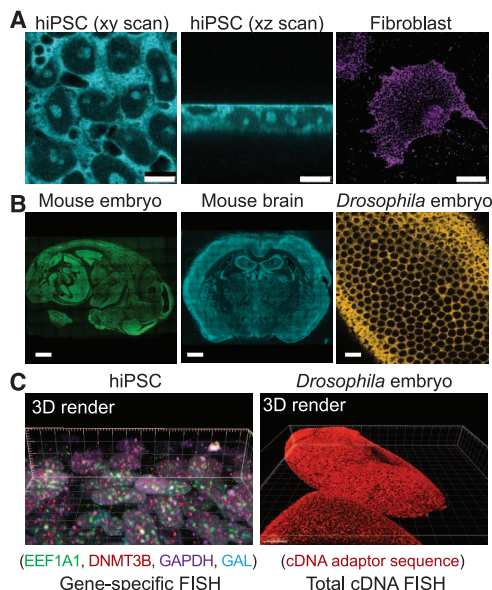


Fig. 1. Construction of 3D RNA-seq libraries in situ. After RT using random hexamers with an adapter sequence in fixed cells, the cDNA is amplified and cross-linked in situ. (A) A fluorescent probe is hybridized to the adapter sequence and imaged by confocal microscopy in human iPS cells (hiPSC) (scale bar: 10 μ m) and fibroblasts (scale bar: 25 μ m). (B) FISSEQ can localize the total RNA transcriptome in mouse embryo and adult brain sections (scale bar: 1 mm) and whole-mount *Drosophila* embryos (scale bar: 5 μ m), although we have not sequenced these samples. (C) 3D rendering of gene-specific or adapter-specific probes hybridized to cDNA amplicons. FISH, fluorescence in situ hybridization.

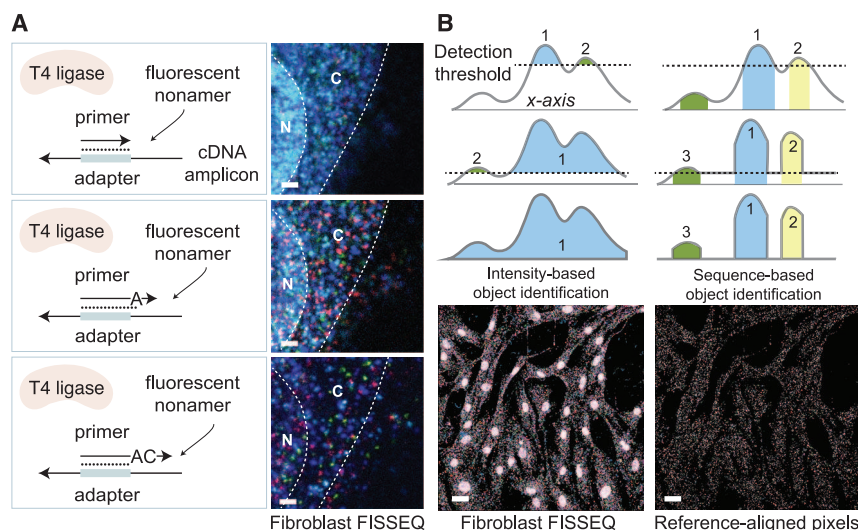


Fig. 2. Overcoming resolution limitations and enhancing the signal-to-noise ratio. (A) Ligation of fluorescent oligonucleotides occurs when the sequencing primer ends are perfectly complementary to the template. Extending sequencing primers by one or more bases, one can randomly sample amplicons at 1/4th, 1/16th, and 1/256th of the original density in fibroblasts (scale bar: 5 μ m). N, nucleus; C, cytoplasm. (B) Rather than using an arbitrary intensity threshold, color sequences at each pixel are used to identify objects. For sequences of L bases, the error rate is approximately $n/4^L$ per pixel, where n is the size of the reference. By removing unaligned pixels, the nuclear background noise is reduced in fibroblasts (scale bar: 20 μ m).

Highly abundant genes in RNA-seq and gene expression arrays were involved in translation and splicing (figs. S11 and S12), whereas such genes were underrepresented in FISSEQ. We examined 12,427 (83.1%) and 2533 (16.9%) amplicons in the cytoplasm and nuclei, respectively, and found

that nuclear RNA was 2.1 [95% confidence interval (CI) 1.9 to 2.3] times more likely to be non-coding ($P < 10^{-16}$), and antisense mRNA was 1.8 [95% CI 1.7 to 2.0] times more likely to be nuclear ($P < 10^{-16}$). We confirmed nuclear enrichment of *MALAT1* and *NEAT1* by comparing their

relative distribution against all RNAs (Fig. 3F) or mitochondrial 16S ribosomal RNA (rRNA) (table S2), whereas mRNA, such as *COL1A1*, *COL1A2*, and *THBS1*, localized to the cytoplasm (table S3). We also examined splicing junctions of *FN1*, given its high read coverage (481 reads over 8.9 kilobases).

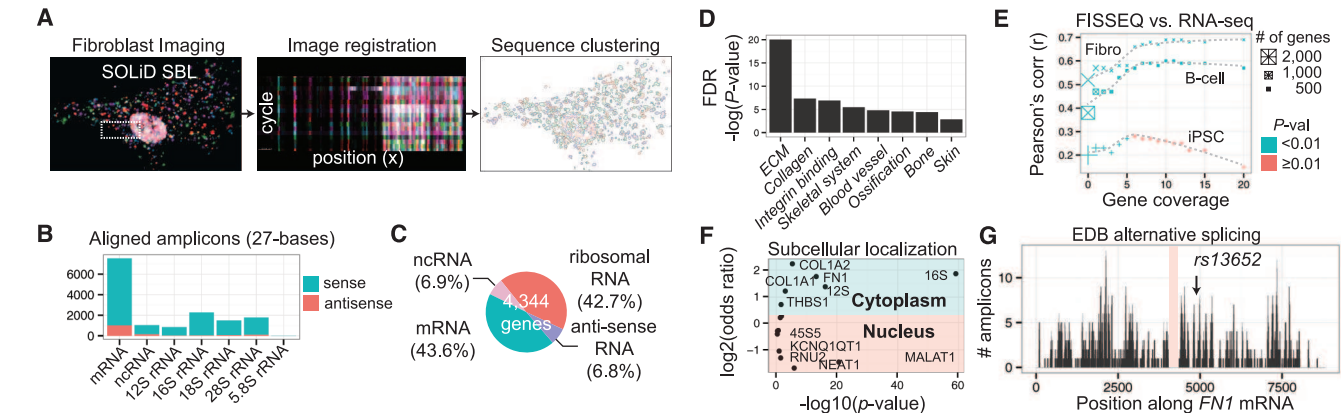
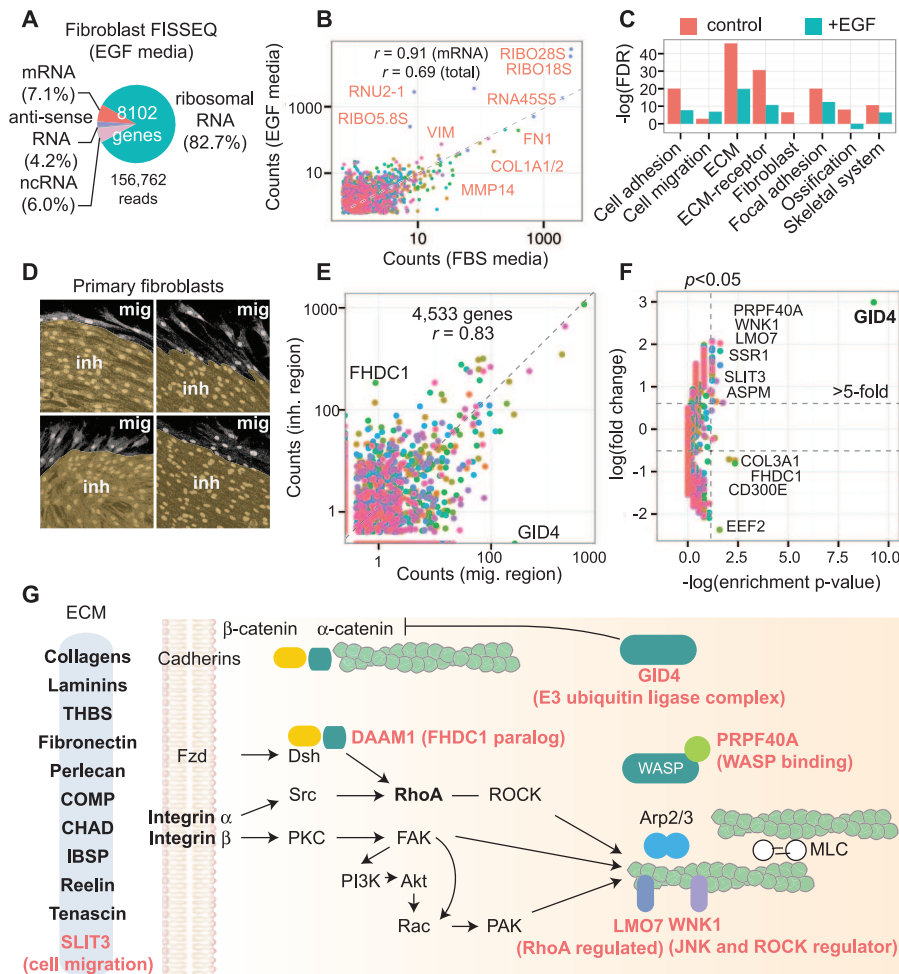


Fig. 3. Whole-transcriptome in situ RNA-seq in primary fibroblasts. (A) From deconvolved confocal images, 27-base reads are aligned to the reference, and alignments are spatially clustered into objects. (B) Of the amplicons, 90.6% align to the annotated (+) strand. (C) mRNA and non-coding RNA make up 43.6% and 6.9% of the amplicons, respectively. (D) GO term clustering for the top 90 ranked genes. (E) FISSEQ of 2710 genes

from fibroblasts compared with RNA-seq for fibroblast, B cell, and iPS cells. Pearson's correlation is plotted as a function of the gene expression level. (F) Subcellular localization enrichment compared to the whole transcriptome distribution. (G) Of the amplicons, 481 map to the *FN1* mRNA, showing an alternatively spliced transcript variant and a single-nucleotide polymorphism (arrow).

Fig. 4. Functional analysis of fibroblasts during simulated wound healing. (A) In EGF medium, rRNA makes up 82.7% of the amplicons. (B) EGF medium 147,610 reads compared with 13,045 reads from FBS medium (different colors denote genes). (C) The top 100 ranked genes from FBS versus EGF FISSEQ clustered for functional annotation. (D) An in vitro wound-healing assay allows cells to migrate (mig) into the wound gap. inh, contact-inhibited cells. The image segments are based on the cell morphology. (E) Comparison of 4533 genes from migrating and contact-inhibited cells. (F) Twelve genes are differentially expressed (Fisher's exact test $P < 0.05$ and >fivefold; 180 genes). (See table S4.) (G) The top 100 genes in fibroblasts are enriched for terms associated with ECM-receptor interaction and focal adhesion kinase complex (bold letters). During cell migration, genes involved in ECM-receptor-cytoskeleton signaling and remodeling are differentially expressed (red letters). THBS, thrombospondin; COMP, cartilage oligomeric matrix protein; CHAD, chondroaderhin; IBSP, integrin-binding sialoprotein; PKC, protein kinase C; FAK, focal adhesion kinase; PI3K, phosphatidylinositol 3-kinase; MLC, myosin light chain; PAK, p21-activated protein kinase; WASP, Wiskott-Aldrich syndrome protein.



FNI has three variable domains referred to as EDA, EDB, and IIICS, which are alternatively spliced (15). We did not observe development-associated EDB, but observed adult tissue-associated EDA and IIICS (Fig. 3G).

We also sequenced primary fibroblasts in situ after simulating a response to injury, obtaining 156,762 reads (>5 pixels), representing 8102 annotated genes (Fig. 4A and fig S13, A to D). Pearson's r was 0.99 and 0.91 between different wound sites and growth conditions, respectively (Fig. 4B and fig. S13, E and F). In medium with epidermal growth factor (EGF), 82.7% of the amplicons were rRNA compared to 42.7% in fetal bovine serum (FBS) medium. When the 100 highest ranked genes were clustered, cells in FBS medium were enriched for fibroblast-associated GO terms, whereas rapidly dividing cells in EGF medium were less fibroblast-like (Fig. 4C) with alternative splicing of *FNI* (fig. S14). In regions containing migrating cells versus contact-inhibited cells, 12 genes showed differences in relative gene expression (Fisher's exact test $P < 0.05$ and >fivefold change) (Fig. 4, D to F, and table S4), eight of which were associated with the extracellular matrix (ECM)–receptor–cytoskeleton interaction, including *GID4*, *FHDC1*, *PRPF40A*, *LMO7*, and *WNK1* (Fig. 4G and table S4).

In summary, we present a platform for transcriptome-wide RNA sequencing in situ and demonstrate imaging and analytic approaches across multiple specimen types and spatial scales. FISSEQ correlates well with RNA-seq, except for genes involved in RNA and protein processing, possibly because some cellular structures or classes of RNA are less accessible to FISSEQ. It

is notable that FISSEQ generates far fewer reads than RNA-seq but predominantly detects genes characterizing cell type and function. If this finding can be generalized, FISSEQ may be used to identify cell types based on gene expression profiles in situ. Using partition sequencing to control the signal density, it may even be possible to combine transcriptome profiling and in situ mutation detection in a high-throughput manner (16–18). Using RNA barcodes from expression vectors, one can label up to 4^N (N = barcode length) cells uniquely, much more than is possible using a combination of fluorescent proteins (19). Similar to next-generation sequencing, we expect advances in read length, sequencing depth and coverage, and library preparation (i.e., fragmentation, rRNA depletion, targeted sequencing). Such advances may lead to improved stratification of diseased tissues in clinical medicine. Although more work remains, our present demonstration is an important first step toward a new era in biology and medicine.

References and Notes

1. G. Diez-Roux *et al.*, *PLoS Biol.* **9**, e1000582 (2011).
2. H. Zeng *et al.*, *Cell* **149**, 483–496 (2012).
3. E. Lécuyer *et al.*, *Cell* **131**, 174–187 (2007).
4. J. Shendure *et al.*, *Science* **309**, 1728–1732 (2005).
5. J. B. Kim *et al.*, *Science* **316**, 1481–1484 (2007).
6. R. D. Mitra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5926–5931 (2003).
7. R. Drmanac *et al.*, *Science* **327**, 78–81 (2010).
8. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, *Science* **280**, 585–590 (1998).
9. J. M. Levisky, S. M. Shenoy, R. C. Pezo, R. H. Singer, *Science* **297**, 836–840 (2002).
10. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, *Nat. Methods* **5**, 877–879 (2008).
11. S. Itzkovitz, A. van Oudenaarden, *Nat. Methods* **8** (suppl.), S12–S19 (2011).

12. M. A. Frohman, M. K. Dush, G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998–9002 (1988).
13. R. J. McAnulty, *Int. J. Biochem. Cell Biol.* **39**, 666–671 (2007).
14. W. Huang, B. T. Sherman, R. A. Lempicki, *Nat. Protoc.* **4**, 44–57 (2009).
15. A. R. Kornblitt *et al.*, *FASEB J.* **10**, 248–257 (1996).
16. R. Ke *et al.*, *Nat. Methods* **10**, 857–860 (2013).
17. K. Zhang *et al.*, *Nat. Methods* **6**, 613–618 (2009).
18. J. H. Lee *et al.*, *PLOS Genet.* **5**, e1000718 (2009).
19. J. Livet *et al.*, *Nature* **450**, 56–62 (2007).

Acknowledgments: Data can be downloaded from http://arep.med.harvard.edu/FISSEQ_Science_2014/ and Gene Expression Omnibus (gene expression arrays: GSM313643, GSM313646, and GSM313657; RNA-seq: GSE54733). We thank S. Kosuri, K. Zhang, and M. Nilsson for discussions; A. DePace for *Drosophila* embryos; and I. Bachelet for antibody conjugation. Funded by NIH Centers of Excellence in Genomic Sciences grant P50 HG005550. J.H.L. and co-workers were funded by the National Heart, Lung, and Blood Institute, NIH, grant RC2HL102815; the Allen Institute for Brain Science, and the National Institute of Mental Health, NIH, grant MH098977. E.R.D. was funded by NIH grant GM080177 and NSF Graduate Research Fellowship Program grant DGE1144152. A.H.M. was funded by the Hertz Foundation. Potential conflicts of interests for G.M.C. are listed on <http://arep.med.harvard.edu/gmc/tech.html>. J.H.L., E.R.D., R.T., and G.M.C. are authors on a patent application from the Wyss Institute that covers the method of generating three-dimensional nucleic acid-containing matrix.

Supplementary Materials

www.sciencemag.org/content/343/6177/1360/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S4
References
Movies S1 to S6

26 December 2013; accepted 13 February 2014
Published online 27 February 2014;
10.1126/science.1250212

Structure of the Mitochondrial Translocator Protein in Complex with a Diagnostic Ligand

Łukasz Jaremko,^{1,2,*} Mariusz Jaremko,^{1,*} Karin Giller,¹ Stefan Becker,^{1†} Markus Zweckstetter^{1,2,3†}

The 18-kilodalton translocator protein TSPO is found in mitochondrial membranes and mediates the import of cholesterol and porphyrins into mitochondria. In line with the role of TSPO in mitochondrial function, TSPO ligands are used for a variety of diagnostic and therapeutic applications in animals and humans. We present the three-dimensional high-resolution structure of mammalian TSPO reconstituted in detergent micelles in complex with its high-affinity ligand PK11195. The TSPO-PK11195 structure is described by a tight bundle of five transmembrane α helices that form a hydrophobic pocket accepting PK11195. Ligand-induced stabilization of the structure of TSPO suggests a molecular mechanism for the stimulation of cholesterol transport into mitochondria.

The 18-kD translocator protein TSPO is preferentially expressed in mitochondrial membranes of steroidogenic tissues, and its gene family is present in almost all organisms (1–3). TSPO was first described as a peripheral benzodiazepine receptor, a secondary receptor for diazepam (1, 4). TSPO was subsequently found to be responsible for the transport of cho-

lesterol into mitochondria, thereby influencing the synthesis of neurosteroids (1, 5). TSPO also plays an important role in apoptosis (6) and stress adaptation (7). Expression of TSPO is strongly up-regulated in areas of brain injury and in neuro-inflammatory conditions including Alzheimer's and Parkinson's diseases (2). TSPO is located at contact sites between the outer and inner mitochon-

drial membrane, and was suggested to be part of (together with cyclophilin D, the voltage-dependent anion channel, and the adenine nucleotide translocator) the mitochondrial permeability transition pore (8).

TSPO ligands have potential diagnostic and therapeutic applications such as attenuation of cancer cell proliferation (6) and neuroprotective effects (2). TSPO ligands such as XBD-173 might also be useful in treating anxiety with reduced side effects relative to traditional benzodiazepine-related drugs (9). The best-characterized ligand of TSPO is 1-(2-chlorophenyl)-*N*-methyl-*N*-(1-methylpropyl)-3-isoquinolinecarboxamide (PK11195), which binds to TSPO with nanomolar affinity in many species (10–13). PK11195 is used as a biomarker in positron emission tomography to visualize brain inflammation in patients with neuronal damage (2, 10). Moreover,

¹Max-Planck-Institut für Biophysikalische Chemie, 37077 Göttingen, Germany. ²Deutsches Zentrum für Neurodegenerative Erkrankungen, 37077 Göttingen, Germany. ³Center for Nanoscale Microscopy and Molecular Physiology of the Brain, University Medical Center, 37073 Göttingen, Germany.

*These authors contributed equally to this work.

†Corresponding author. E-mail: markus.zweckstetter@dzne.de (M.Z.); sabe@nmr.mpiibpc.mpg.de (S.B.)



Supplementary Materials for

Highly Multiplexed Subcellular RNA Sequencing in Situ

Je Hyuk Lee,* Evan R. Daugharthy, Jonathan Scheiman, Reza Kalhor, Thomas C. Ferrante, Joyce L. Yang, Richard Terry, Sauveur S. F. Jeanty, Chao Li, Ryoji Amamoto, Derek T. Peters, Brian M. Turczyk, Adam H. Marblestone, Samuel A. Inverso, Amy Bernard, Prashant Mali, Xavier Rios, John Aach, George M. Church*

*Corresponding author. E-mail: jehyuklee@mac.com (J.H.L.);
gchurch@genetics.med.harvard.edu (G.M.C.)

Published 27 February 2014 on *Science Express*
DOI: 10.1126/science.1250212

This PDF file includes

Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S4
Movie legends S1 to S6
References

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.1250212/DC1)

Movies S1 to S6

Correction: The correct Fig. S1 has replaced Fig. 1 from the report.

Materials and Methods:

Cell Lines. An openly consented and IRB-approved Personal Genome Project (PGP) iPS cell line can be obtained from Coriell (GM23338). The donor primary fibroblast is GM23248. The doxycycline-inducible mCherry HeLa cell line is constructed by combining the TetON-3G system (Clontech) and the Piggybac transposon (Systems Bioscience) into a single doxycycline-inducible transposable vector.

Cell Culture. iPS cells are grown in mTeSR1 media (Stem Cell Technologies) on ES cell-qualified Matrigel-coated plates (EMD Millipore). Primary fibroblasts are grown in 10% fetal bovine serum D-MEM or 15% newborn calf serum D-MEM/F12 with 10 ng/ml human EGF, 1× L-glutamine, pen/strep, and nonessential amino acid (Life Technologies). HeLa cells are grown in 10% FBS DMEM with GlutaMAX, sodium pyruvate, pen/strep, and 10 ug/mL puromycin (Life Technologies). All cells are grown on glass bottom Petri dishes (Mattek dish part No. P35GC-1.5-14-C).

Wound closure assay. Human primary fibroblasts are grown to 100 percent confluency in EGF-supplemented media on a glass bottom Petri dish. Using the corner of a glass cover slip, multiple scratches (~200-um wide) are made on the cell monolayer. Eight to twelve hours after the first scratch, the cells are fixed for analysis.

Amplicon stretching on glass assay. RCA amplicons are generated *in vitro* using a synthetic DNA template under FISSEQ RCA conditions. Aminosilane-treated glass is prepared by incubating cleaned #1.5 glass coverslips for 10 minutes in 1% (3-Aminopropyl)triethoxysilane (Sigma) in acetone. RCA amplicons are bound to the glass in PBS and fixed under FISSEQ fixation conditions. Amplicons were counted using ImageJ.

Cell fixation and permeabilization. Cells are fixed with 4% formaldehyde (Sigma) for 15 minutes. iPS cells are permeabilized using 0.25% Triton-X100 (Calbiochem) for 15 minutes, and primary fibroblasts are permeabilized using 70% ethanol for 5 minutes. Samples are then incubated with 0.1 N hydrochloric acid for 2 minutes. For tissue sections, 0.01% of pepsin (Roche) in 0.1 N hydrochloric acid is used for permeabilization, followed by three PBS washes to neutralize pepsin.

In situ reverse transcription and amplification. A 200 uL mixture containing 4,000 U M-MuLV reverse transcriptase (Enzymatics), 250 uM dNTP (Enzymatics), 40 uM aminoallyl dUTP (Anaspec), 50 U RNase inhibitor (Enzymatics), and 100 pmol tagged random hexamers (/5Phos/TCTCGGGAACGCTGAAGANNNNNN), prepared on ice is added to cells at 25°C for 10 minutes. The concentration of aminoallyl dUTP can vary depending the cell type and the application. Generally, a high incorporation rate of aminoallyl dUTP results in better cross-linking and reduced cDNA diffusion but a lower amplicon density. The sample is then incubated overnight in a humidified 37°C chamber. The sample is washed using 1x PBS and cross-linked using BS(PEG)9 (Thermo Scientific), diluted to 50 mM in PBS, for 1 hour at 25°C. 1 M Tris (G Biosciences) is added to quench the reaction for 30 minutes at 25°C. A mixture of DNase-free RNases (Roche Diagnostics) and RNase H (Enzymatics) is added to degrade residual RNA for 1 hour at 37°C. A 100 uL circularization reaction mixture (1x reaction buffer, 2.5 mM MnCl₂, 1 M Betaine and 5 uL CircLigase II from Illumina/Epicentre) is then added to the sample well and incubated at 60°C for 2 hours. After circularization, the sample is washed using H₂O and incubated with a 200 uL mixture containing 0.1 uM RCA primer (TCTTCAGCGTTCCCGA*G*A from IDT) in 2x SSC and 30% formamide for 15 minutes at 60°C. The sample is washed using 2x SSC, and a 200 uL amplification mixture containing 500 U

Phi29 DNA polymerase (Enzymatics), 250 uM dNTP, 40 uM, and aminoallyl dUTP is added. The sample is incubated in a dry 30°C chamber overnight and cross-linked using BS(PEG)9 diluted to 50 mM in PBS for 1 hour at 25°C. After a rinse with PBS, 1 M Tris is added to quench the reaction for 30 minutes. At this point, the sample can be stored in nuclease-free 1x PBS at 4°C.

SOLiD dinucleotide sequencing. We designed five sequencing primers specific to our universal adaptor (N, N-1, N-2, N-3, N-4, where N- x is recessed at the 5' end by x -nt). These sequencing primers are annealed to the sample sequentially, and k ligation reactions are performed for each primer ($k+1$ ligation reactions for primers N-2, N-3, and N-4). Each sequencing primer is annealed to the sample at 2.5 uM in 200 uL 80°C 5X SASC (0.75 M sodium acetate and 75 mM tri-sodium citrate, pH 7.5), incubating for 10 minutes at 25°C. The sample is washed twice for one minute each with 1 mL 1x Instrument Buffer (SOLiD Instrument Buffer Kit, Applied Biosystems Cat# 4389784). 200 uL sequencing mix is freshly prepared on ice using 165 uL nuclease-free H₂O, 20 uL T4 DNA ligase buffer, 10 uL T4 DNA Ligase (Enzymatics), and 5 uL SOLiD sequencing oligos (the dark purple tubes from the SOLiD ToP Sequencing Kit Fragment Library F3 Tag MM50 Cat# 4449388). After aspirating the Instrument Buffer, the sequencing mix is added to the sample and incubated at 25°C for 45 minutes. The sequencing mix is aspirated, and the sample is washed with 1x Instrument Buffer (four 1 mL washes for 5 minutes each). Imaging is done in 1 mL 1x Instrument Buffer. After aspirating the Instrument Buffer, the fluorophore is cleaved to allow for subsequent ligation. The sample is incubated twice for 5 minutes each in 200 uL 1x Cleave Solution 1 (SOLiD ToP Instrument Buffer Kit Component 4406489), followed by two incubations for five minutes each in 200 uL 1X Cleave Mix 2.1 (SOLiD ToP Instrument Buffer Kit Component 4445677, prepared fresh with 106.7 uL Cleave

Solution 2.1 Part 1 and 293.3 uL Cleave Solution Part 2). After the second incubation with Cleave Mix 2.1, the sample is washed three times for 5 minutes each with 1x Instrument Buffer. After repeating the cyclic ligation process k (or $k+1$) times, the ligated strands are stripped by four 5 minute washes in 80°C strip buffer (80% formamide, 0.01% Triton-X100). Another sequencing primer is annealed, and the cyclic ligation process is repeated.

Partition Sequencing. Fundamentally, any set of orthogonal primers can selectively sequence the amplicons for improved spatial resolution, provided that RT primers contain complementary adapter sequences. Our method (Fig. 2) uses sequencing primers that extend into the cDNA sequence by several nucleotides. The orthogonality defined by single nucleotide differences is based on the high specificity of T4 DNA ligase near the ligation junction (<6 bases away). With the template 3'-Adapter-NNN...-5', the sequencing primer can be one of 5'-Adapter (reverse complement)-A/G/C/T-3'. N is degenerate, so ligation occurs on $1/4^{\text{th}}$ of the amplicons that start with a defined base (T/C/G/A). The sequencing primer can also be one of 5'-Adapter (reverse complement)-(A/G/C/T)₂-3' or 5'-Adapter (reverse complement)-(A/G/C/T)₃-3' for detecting $1/16^{\text{th}}$ or $1/256^{\text{th}}$ of the amplicons. Sequencing reads from multiple primers can be combined, increasing spatial resolution, sequencing time, and cost. Partition sequencing can be useful for short barcode sequencing and *in situ* quantitation by effective serial dilution. Because SOLiD sequencing uses recessed primers itself, partition sequencing using SOLiD requires synthesis of specific bridge oligonucleotides. Instead, we use sequencing-by-ligation detailed in http://openwetware.org/wiki/Church_Lab (4).

Image acquisition. The sample is firmly clamped to the stage of a confocal microscope. Imaging is done on a Zeiss Axio Observer with LSM 710 scanning laser confocal system, using the following excitation and emission profile. FITC: 25 mW Argon laser (458/488/514 nm) and 490-

560 nm emission filter. Cy3: 20 mW DPSS laser (561 nm) and 563-593 nm emission filter.

Texas Red: 2 mW HeNe laser (594 nm) and 597-647 nm emission filter. Cy5: 5 mW HeNe laser (633 nm) and 637-758 nm emission filter. Typically, we choose pixel resolution and optical slice thickness close to the Nyquist sampling rate (<http://www.svi.nl/NyquistCalculator>).

Image processing. The raw images are first deconvolved using Huygens Professional 4.3 running on a Windows 7 Intel Xeon workstation (typically less than five iterations) to increase the signal to noise ratio and reduce dephasing over sequencing cycles. For relatively flat cells such as fibroblasts, maximum intensity projection (MIP) is created from multiple optical slices in order to compensate for z-drift over time and to reduce the computational time required for image alignment. The separate images from different sequencing cycles form a time series (up to 30 time points) containing four colors. Using custom Matlab and Python scripts, individual images are corrected for chromatic shifts and registered locally (using a brick-based algorithm with 100 bricks per image). Software can be downloaded from www.arep.med.harvard.edu.

Base calling and spatial clustering. In order to identify the amplicon sequences, as well as their alignment to the reference sequence library, amplicons are treated as a set of spatially connected pixels with the similar color transitions over the sequencing run. Pixels at the amplicon border can have missing base calls due to reduced quality or misalignment, but short-read aligners can tolerate small mismatches. To cluster pixels with a shared alignment to the reference into objects, we first import the Tiff images and use the vector of fluorescent intensities to calculate a component vector for each pixel: $C_x = I_x / (I_1^2 + I_2^2 + I_3^2 + I_4^2)^{1/2}$, where I_x indicates the intensity for channel x , and C_x represents the component value for channel x . The base call is determined by the largest component value at each pixel. Pixels without a largest component value are masked as not having a base call. The quality of the base call is determined as the geometric

distance between the unit component vector and the unit vector in the direction of the base call.

Another measure of quality, used for Satay plots, is the angle between the unit component vector and the unit vector in the direction of the base call, calculated as $\arccos(1-\theta)$, where $\theta = C_x$ for the component value corresponding to the base call. Reads are generated for all pixels with less than 3-6 missing bases per read. Reads are then written to the csfasta format (0 = FAM, 1 = CY3, 2 = TXR, 3 = CY5, '.' = no base call) for alignment using the SOLiD color space-compatible short-read aligner Bowtie v1. The alignment output file from Bowtie is loaded back into the custom Python program, where alignments are re-assigned to the corresponding pixels using the read ID. For each alignment class (defined as having a particular reference sequence, strand, and position of alignment), all pixels with mapped reads are spatially clustered with a user-specified kernel. (We defined connectedness as being separated by no more than one pixel.) Clusters are redefined as amplicons occupying a set of pixels and having various statistics such as centroid, consensus mismatches (>50% of the pixels in the amplicon share a particular Bowtie mismatch), mean per-base quality, etc. Amplicons can have additional alignment classes if more than 1/2 of the pixels are shared between an existing amplicon and a new cluster generated from a different alignment class. This allows an amplicon to have multiple alignments, such as when a read aligns identically to several variants of a transcript. For each amplicon, a best alignment class is defined as the set of alignments with the highest strata of mean alignment quality over all bases and pixels. The best alignment class is exclusively used for filtering and in downstream analysis. Amplicons are then filtered by size and summary statistics are written to file. Per-base sequencing error rate is calculated as the total number of consensus mismatches at each base divided by the total number of consensus reads (e.g. Base 1 Error Rate = (# Amplicons w/

Consensus Mismatch at Base 1) / (Total # Amplicons). The mean and median per-base error statistics are generated using these values.

Bowtie read alignment. Reads are aligned separately to several references to allow granularity in the Bowtie settings: Human RefSeq RNA (containing all RefSeq NM and NR class annotations), human ribosomal RNA reference (including mitochondrial rRNAs not found in the human RefSeq reference), and human tRNA reference. For the first experiment, the 27 sequenced bases correspond to nucleotides 1-25, plus nucleotides 28 and 29, after the sequencing primer. Since base calls are missing for nucleotides 26 and 27, we allow up to 3 mismatches of any base call quality value in the seed region (Bowtie flag: -n 3 -l 15 -e 240). In both cases, all alignments in the best strata (Bowtie flags: --best --strata) are reported for reads with less than 20 alignments (Bowtie flag: -m 20). All statistics are calculated using a read length value reflecting the number of bases actually sequenced. In plots and figures, any statistics for the missing bases are excluded but can be understood to be zero, since no sequencing data are acquired for these bases.

Expression array and RNA-Seq. The total RNA is isolated from ~500,000 PGP1 primary fibroblasts, immortalized B-cells, and iPS cells (RNeasy, Qiagen) for BeadChip HuRef-8 v3 expression arrays (Illumina) and RNA-seq. For RNA-seq, the cDNA library is generated using random hexamers and poly dT reverse transcription primers for linear displacement amplification (Ovation RNA-seq System, NuGEN). After size selection, the sequencing library is prepared using SPRIworks Fragment Library System (Beckman Coulter) and sequenced on HiSeq2000 (Illumina) for 75-base paired-end reads (Partners HealthCare Center for Personalized Genetic Medicine, Harvard). We use fastq-mcf, BWA, and eXpress 1.4.0 for read processing, alignment (Human RNA RefSeq), and gene quantification. The total number of paired-end reads

mapped to the transcriptome from primary fibroblasts, B-cells, and iPS cells are 27.2 million, 14.1 million, and 12.4 million, respectively.

Subcellular transcript localization & differential expression. Since the nuclei are brighter in our images than the rest of the sample, we use the MATLAB image processing toolbox and manual annotation to generate nuclear masks, and determine whether the centroid of each amplicon is located inside or outside of the nuclear mask. For each gene, the cytoplasmic to nuclear ratio is compared to that of other genes with a known subcellular localization profile using Fisher's exact test in R to generate *p*-values, odds ratios, and 95% confidence intervals. In most cases, we are comparing <200 genes, so our *p*-values are not multiple hypotheses corrected. For differential expression we used genes with >5 total observations to calculate Fisher's exact test *p*-values, odds ratios, and 95% confidence intervals.

Data visualization & plots. We use Bitplane Imaris 7.6 for visualizing 3D images during the sequencing run and for creating movies. Gene expression data visualization is done using ggplot2 in R. In cases where multiple genes have the same expression counts, we add random noise ('jitter plot') to avoid over-plotting and use multiple colors to denote different genes.

Statistics. For gene ontology analysis, a default setting in DAVID 6.7 is used. For comparing FISSEQ to RNA-seq, Pearson's correlation is generated between the datasets after applying various minimum and maximum count thresholds to the FISSEQ dataset. Increasing the minimum count threshold increases the correlation between FISSEQ and RNA-seq. Reducing the maximum count threshold initially improves the correlation by excluding *FNI* (>500 counts; most genes are under 200 counts). A further reduction of the maximum count threshold leads to a lower correlation. Thus, we treat *FNI* as an outlier for expression level analysis.

Supplementary Text

Additional author notes: J.H.L. and E.R.D. developed the method, performed imaging and sequencing and analyzed the data. J.S. prepared and validated the FISSEQ libraries used in this experiment. R.K. designed and performed the mCherry quantification experiment. J.L.Y. and R.A. performed additional *in situ* RNA-seq library construction. S.A.I., R.T., C.L., D.T.P., B.M.T. assisted with automation. P.M. and X.R. generated the fluorescently tagged cell lines, and S.S.F.J. provided the human iPS cells. A.B. generated the mouse brain sections, and A.H.M. generated the AFM and the TEM images. T.C.F. assisted in developing high-resolution imaging for FISSEQ, and D.P. assisted on partition sequencing. J.A. advised on the study design. J.H.L., E.R.D. and G.M.C. conceived, designed and wrote the paper. J.H.L. and E.R.D. contributed equally to this manuscript.

Fig. S1. Fluorescent *in situ* sequencing (FISSEQ) library construction. **(1)** A tagged random hexamer primer is used to prime M-MuLV reverse transcriptase to generate aminoallyl dUTP-modified cDNA fragments in fixed cells or tissues. **(2)** BS(PEG)9 permanently cross-links the modified cDNA and the cellular protein matrix. **(3)** After cDNA circularization, **(4)** Phi29 DNA polymerase generates cDNA amplicons **(5)** cross-linked to form **(6)** the 3D *in situ* RNA sequencing library within the cell.

Fig. S2. Spatial stabilization of the RNA-seq library *in situ*. **(A)** A short RT primer (24-mer) cannot circularize and amplify *in situ*, unlike the cDNA (>40-bases). **(B)** Aminoallyl dUTP can be incorporated during RT, providing the reactive group for BS(PEG)9. Without cross-linking, formamide removes much of the cDNA in iPS cells, reducing the number of amplicons (bar: 10 μ m). **(C)** With cross-linking, a formamide wash does not reduce the number of amplicons in iPS cells. **(D)** A circular template covalently linked to the protein matrix amplifies efficiently *in situ*. Here an anti-actin antibody is conjugated to the synthetic DNA, purified using ion exchange chromatography, and used to label primary fibroblasts prior to RCA (bar: 10 μ m). **(E)** RCA amplicons on an aminosilane-treated glass surface are washed aggressively, stretching >95% of unfixed amplicons vs. ~25% of fixed amplicons (bar: 3 μ m).

Fig. S3. Improving the amplicon density *in situ*. **(A)** The cDNA yield is typically low when random hexamers are used to prime reverse transcription, even in solution. **(B)** The resulting single-stranded cDNA fragments do not circularize efficiently (upward mobility shift) unless the residual RNA is degraded using RNase A (* denotes exonuclease-contamination). **(C)** Using real-time qPCR, the amount of circular GAPDH cDNA was estimated in the random hexamer-primed cDNA library with and without 1 hour CircLigase (CL) or RNase A (RA) after

Exonuclease I treatment. **(D)** Without RNases, the density of cDNA amplicons *in situ* is variable and typically low in iPS cells.

Fig. S4. Characterization of the cDNA amplicons. **(A)** A standard RCA amplicon generated in solution is a negatively charged DNA molecule containing a large number of the template sequence (AFM bar: 1 μ m, TEM bar: 100 nm). **(B)** BS(PEG)9 renders the surrounding area less prone to non-specific probe binding. **(C)** Compared to single molecule RNA FISH (20 probes to human GAPDH), FISSEQ amplicons retain their signal-to-ratio even after a prolonged exposure to an excitation laser. **(D)** The amplicons can be re-hybridized 60 or more cycles without a reduction in the signal-to-noise ratio. **(E)** The average centroid drift during sequencing in iPS cells. **(F)** An average amplicon in fibroblasts (overlay of 14,960 centroid-aligned amplicons with 27 nt reads). On average, signals from 27 sequencing cycles overlap by $\sim$ 36 pixels (600 nm diameter) using a 20x objective (N.A. 0.75).

Fig. S5. Sequencing reaction cycles and imaging. **(A)** Modified SOLiD sequencing-by-ligation for FISSEQ in a primary fibroblast, and deconvolved images for the first 15 cycles. **(B)** Laser excitation power and gain settings over the thirty imaging cycles.

Fig. S6. The high amplicon density enables visualization of the RNA-rich subcellular compartments in iPS cells. **(A)** In some iPS cell populations, cDNA amplicons are enriched in the cytoplasm and the nucleolus, but not the nucleus. **(B)** In other iPS cell populations, the cDNA amplicons are enriched in the nucleus, counter-staining the nucleolus or condensed chromatin bodies. Amplicon-free speckles are frequently observed in the nucleus and the cytoplasm, potentially indicating regions without RNA or not accessible to FISSEQ.

Fig. S7. Single gene capture and sequencing *in situ* in HeLa cells. **(A)** The doxycycline-inducible mCherry transcripts are reverse transcribed using a series of RT primers along the 5' region in

order to assess the effect of cDNA length on the sensitivity. **(B)** Upon doxycycline treatment, the mCherry cDNA amplicons are detected in a concentration- and time-dependent manner (each point is a replicate experiment of ~300 cells). **(C)** The 3' end of the cDNA corresponds to the transcription start site (TSS), which abuts the 5' end of the adapter sequence when circularized. **(D)** When the amplicons are sequenced *in situ*, 3,967 out of 7,492 (52.4%) amplicons map to the four base window spanning the predicted TSS.

Fig. S8. Imaging and base calling statistics. **(A)** 3D deconvolution enhances the signal-to-noise ratio and decreases dephasing of the base-specific fluorescence over multiple sequencing cycles in a fibroblast (bar: 5 μ m). The high background represents the nucleus. **(B)** The number of amplicons associated with specific fluorescence is plotted as a function of sequencing cycles. **(C)** A Satay plot of amplicon fluorescence from all 27 sequencing cycles. The radius is the intensity of the signal, and the angle is the separation between the signal and the axis of the called base. **(D)** The basecall quality over 27 sequencing cycles. **(E)** The median per-base error rate from the whole transcriptome FISSEQ as a function of cluster size.

Fig. S9. FISSEQ image and data analysis pipeline. Confocal image stacks deconvoluted to reduce dephasing of base calls. Depending on the sample thickness and the amplicon density, the number of z-slices are reduced to minimize the processing time. Images were corrected for chromatic shifts and registered to one another using a block-based algorithm prior to applying sequenced-based quality filters. The remaining pixels are spatially clustered to identify specific sequence-associated amplicons of size >5 pixels.

Fig. S10. Basecall and alignment quality in primary fibroblasts. **(A)** The distribution of the base quality as a function of the amplicon pixel cluster size. **(B)** The centroid of aligned amplicons is plotted as a function of the basecall quality. **(C)** The alignment quality and **(D)** the amplicon size

(or the number of pixels per amplicon) are relatively uniform throughout the whole image and within the cell.

Fig. S11. Comparison of the gene expression data from expression arrays and FISSEQ. **(A)** The top 100 ranked genes from microarrays are shown. **(B)** The top 90 ranked genes from FISSEQ are enriched in genes known to be markers of the fibroblast activity. **(C)** The correlation between the array and FISSEQ dataset among genes with different expression array levels. Pearson's correlation is plotted as a function of the expression array level. The number of genes used for correlation is indicated on the right. **(D)** Compared to the expression array, FISSEQ has a false positive rate of 6% and detects fewer genes related to RNA processing and splicing ($\text{FDR} < 10^{-13}$).

Fig. S12. Comparison of the functional term enrichment between RNA-seq and FISSEQ. Despite a much deeper coverage and read depth, the top 1,000 ranked genes from RNA-seq do not form cell type-specific functional clusters. The top 90 ranked genes from FISSEQ are related to the bone and skin development (Fig. 3D), while the top 1,000 ranked genes from FISSEQ are associated with wound healing and skeletal development.

Fig. S13. Wound healing FISSEQ across five different regions. **(A)** Basecall quality as a function of the amplicon size (in pixels). **(B)** Alignment quality as a function of the amplicon size. **(C)** Sequence complexity as a function of the amplicon size. **(D)** Gene categories as a function of the amplicon size. **(E)** Gene expression comparison between wound regions ($n=658$ genes). Pearson's r is >0.999 for all genes, 0.91 for genes with 10-100 counts, and 0.41 for genes with 0-10 counts. **(F)** Gene expression comparison between slow and fast growing fibroblasts in separate experiments ($n=2,309$ all genes; $n=1,621$ mRNA).

Fig. S14. Analysis of alternative splicing of *FNI* *in situ*. **(A)** Approximately 500 *FNI* reads from the cells grown in FBS media and EGF media are compared. Reads aligned to Exon 25 (EDB)

are enriched in fibroblasts from EGF media. **(B)** Retention of Exon 25 (EDB) is 90-fold higher in EGF media, compared to slower growing fibroblasts in FBS media; however, no statistical enrichment of EDB retention is seen between migrating and contact inhibited fibroblasts. Error bars indicate a 95% confidence interval of the estimated odds ratio (Fisher's exact test).

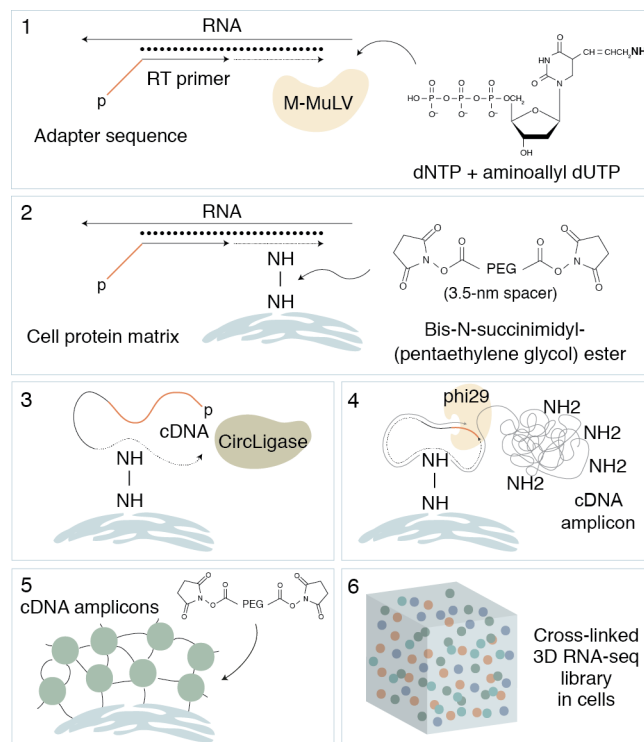


Figure S1

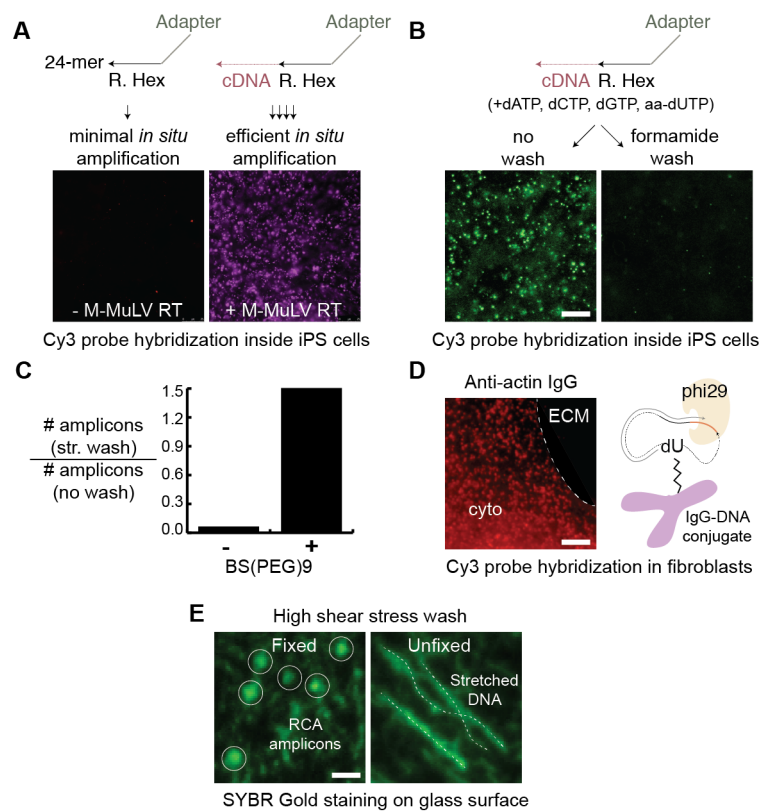


Figure S2

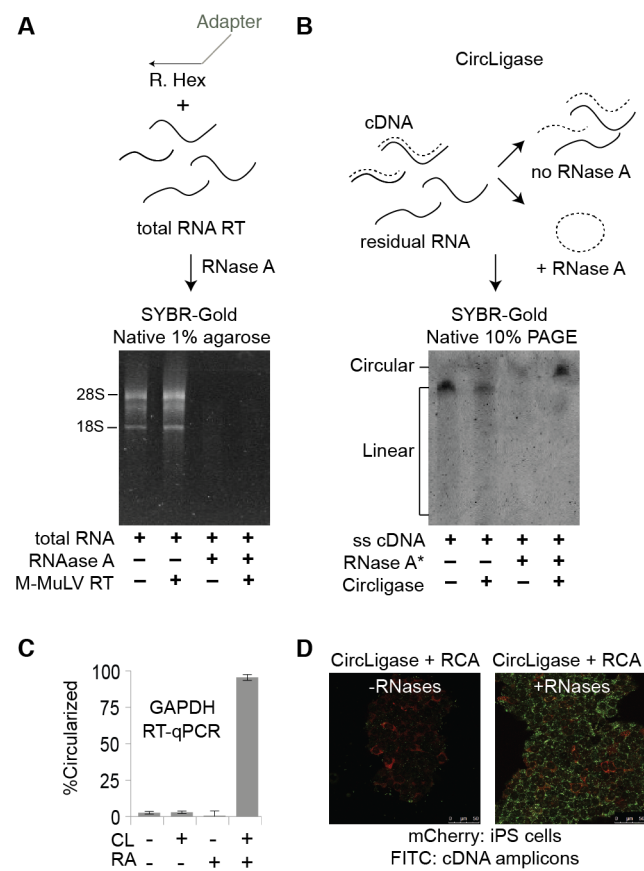


Figure S3

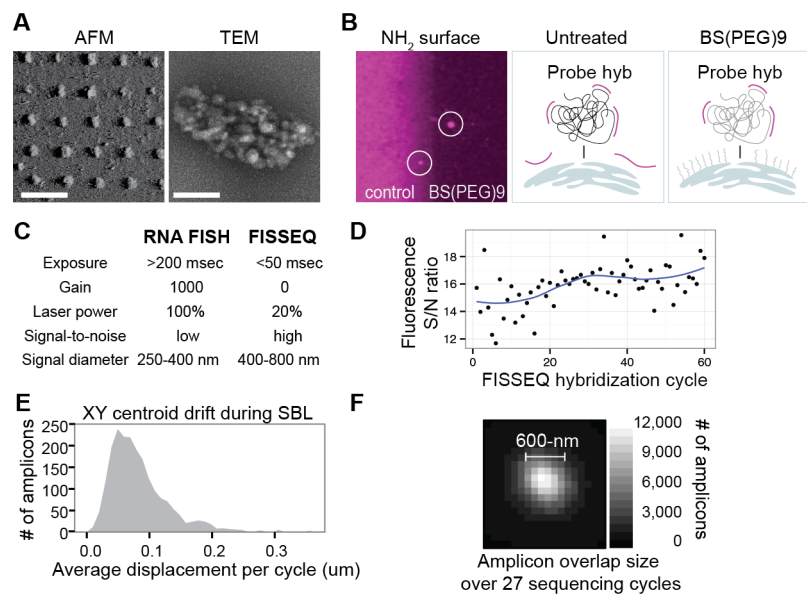


Figure S4

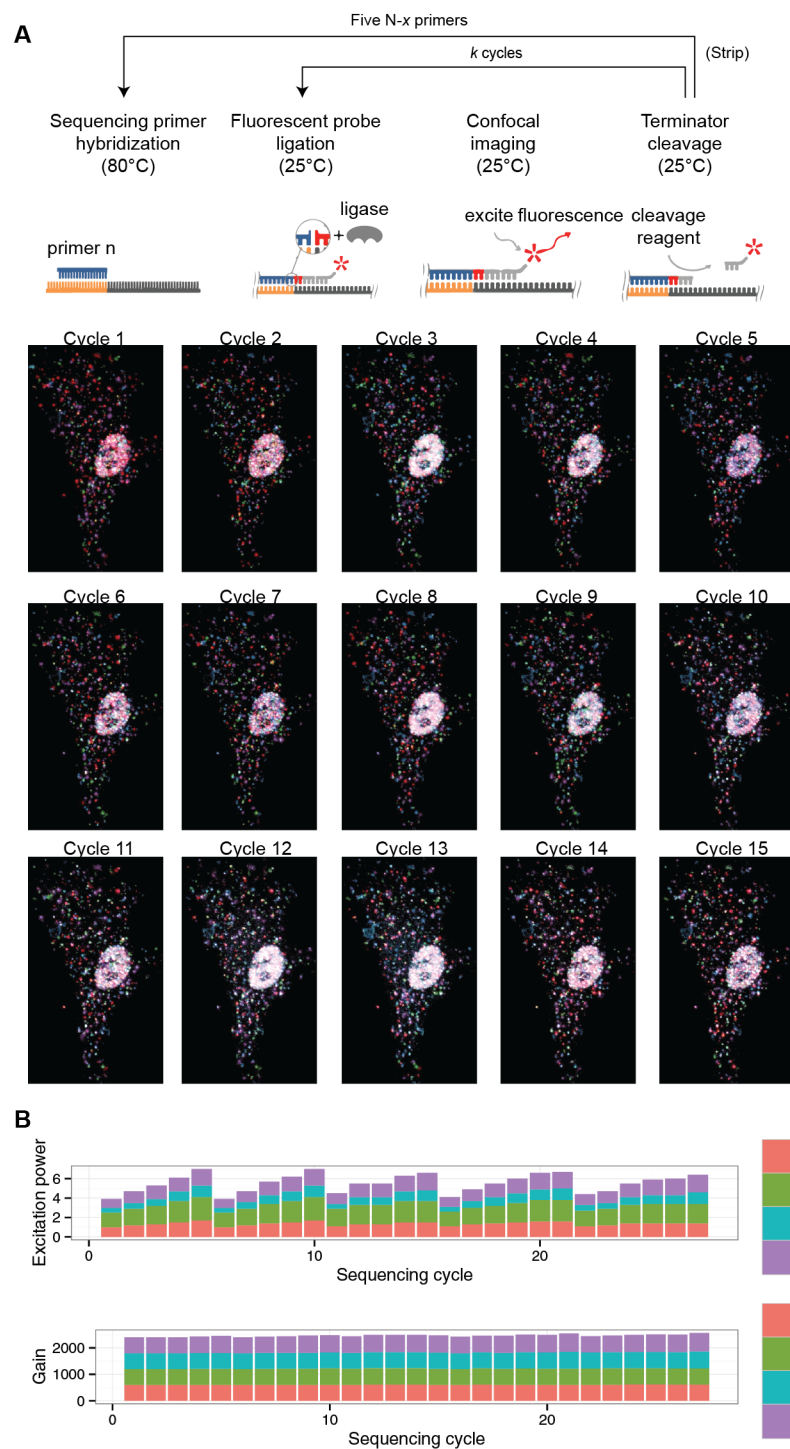


Figure S5

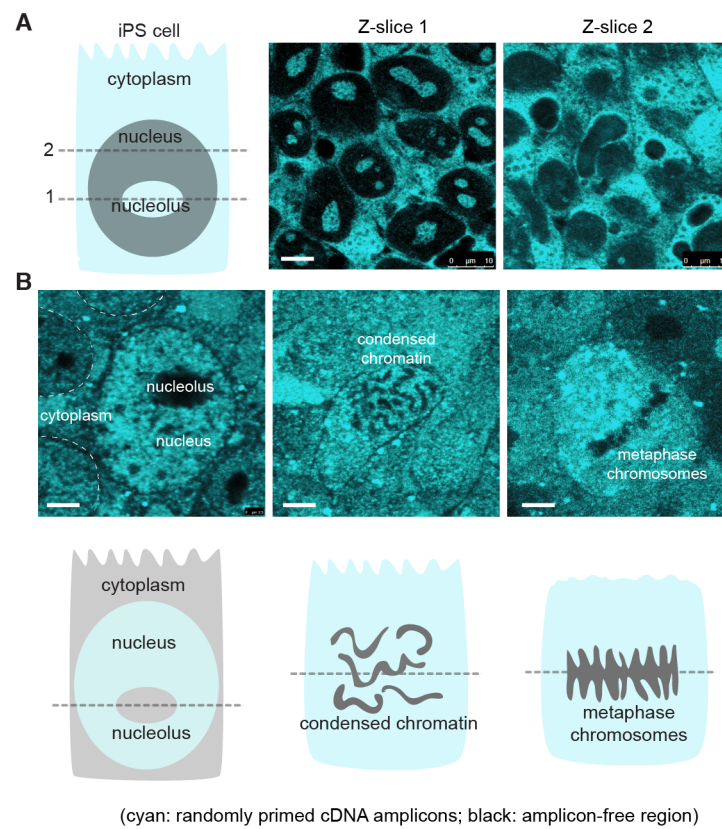


Figure S6

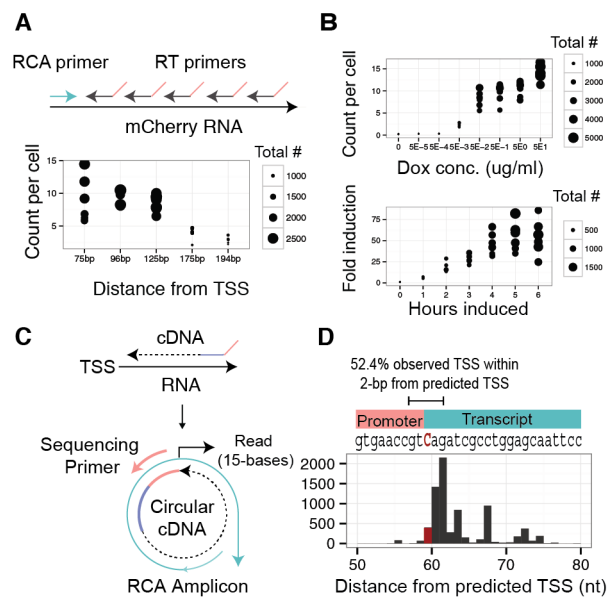


Figure S7

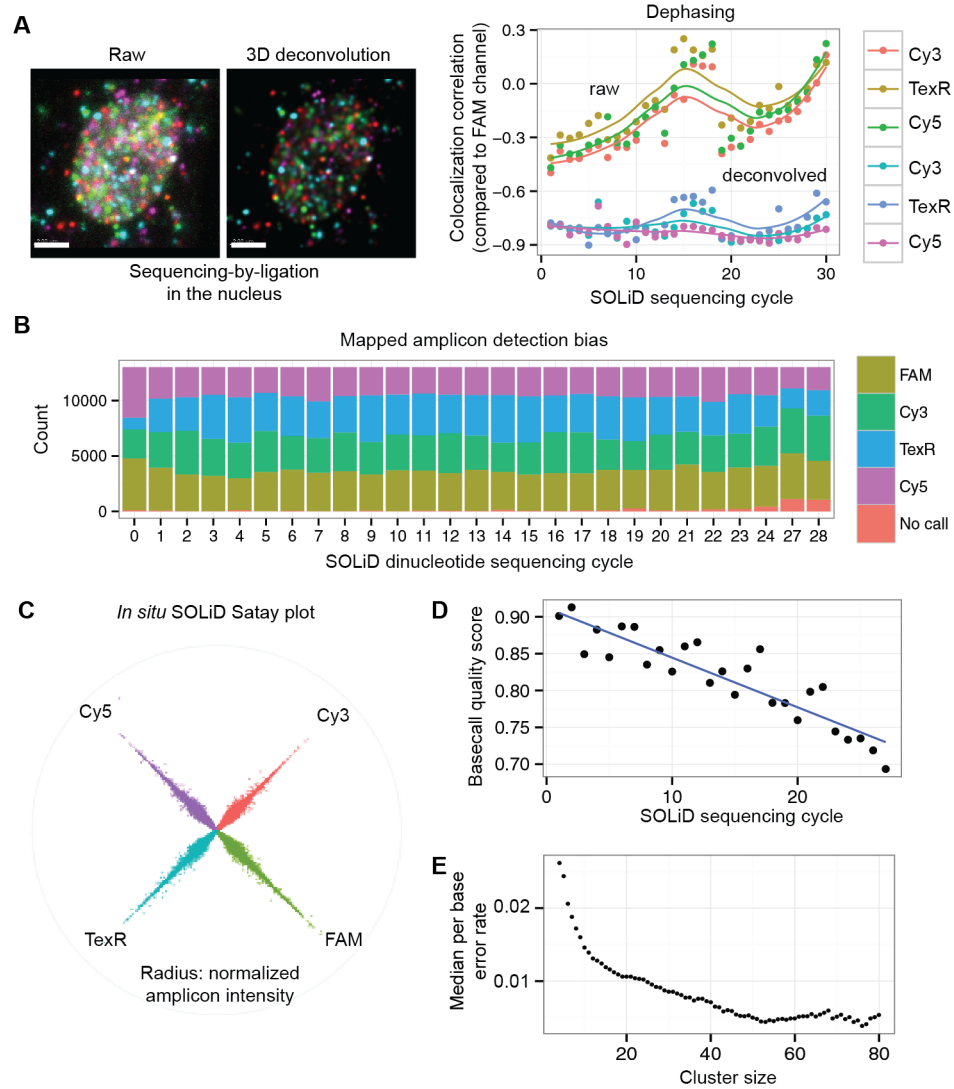


Figure S8

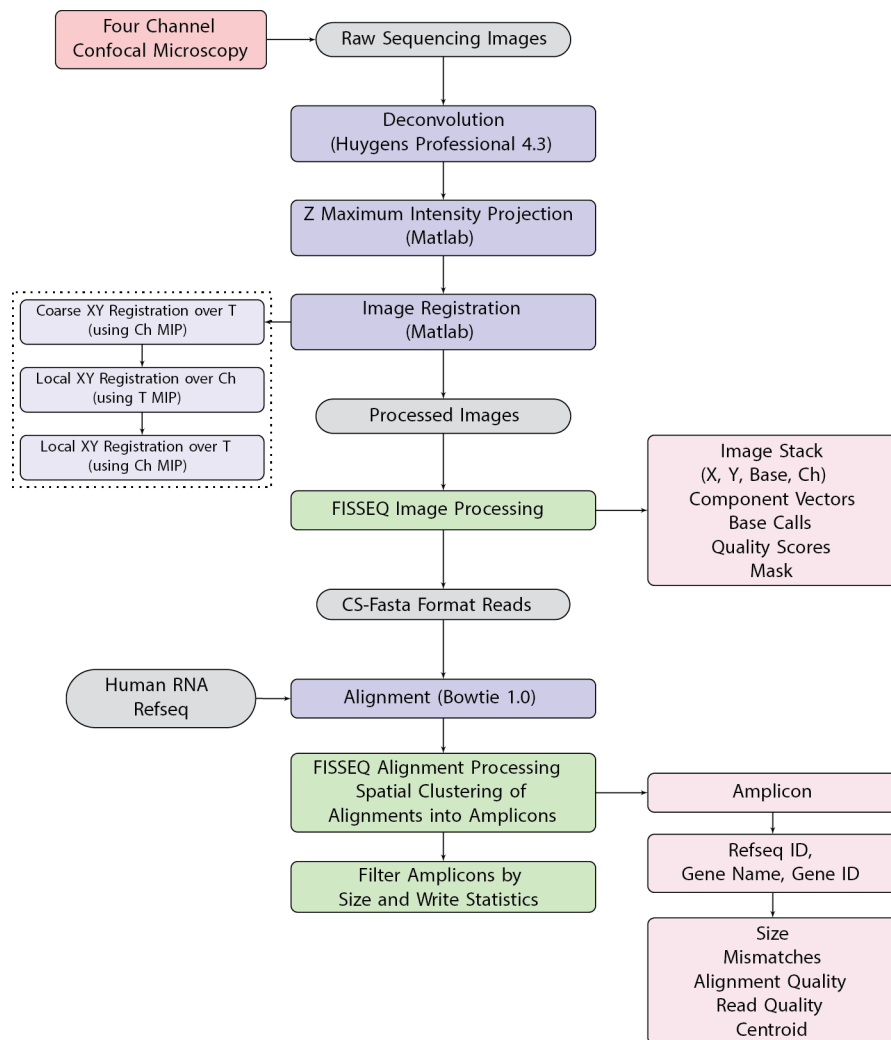


Figure S9

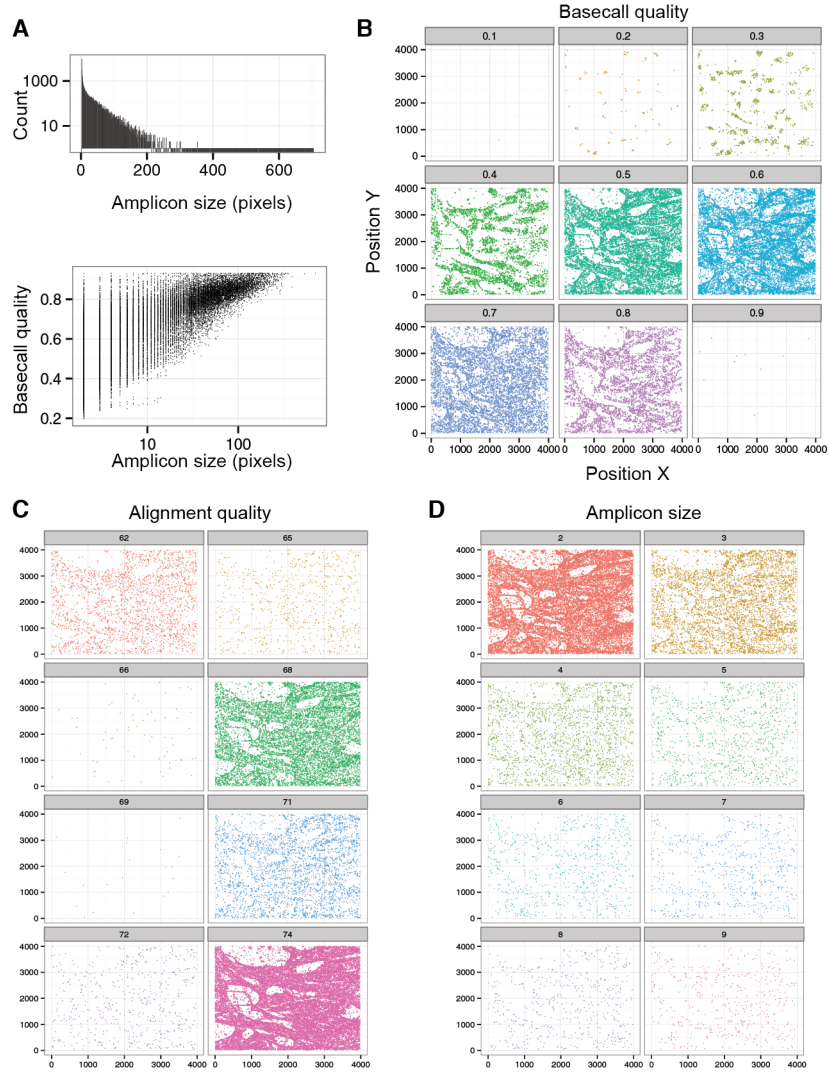


Figure S10

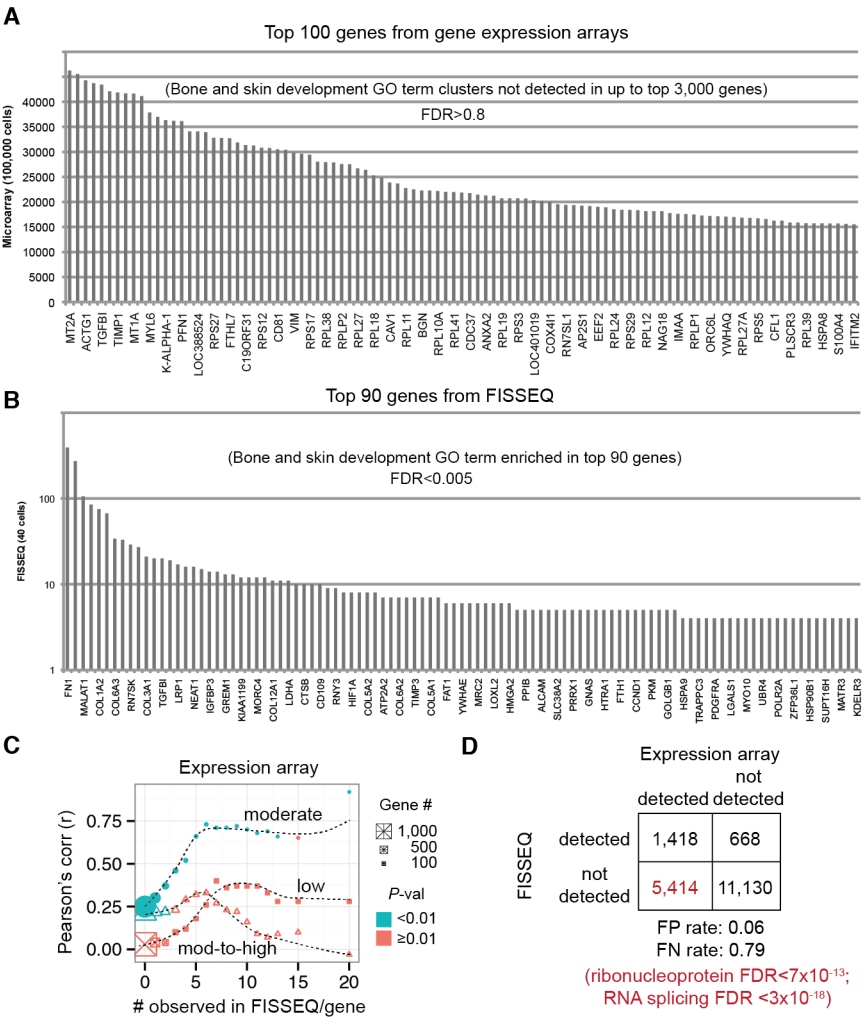


Figure S11

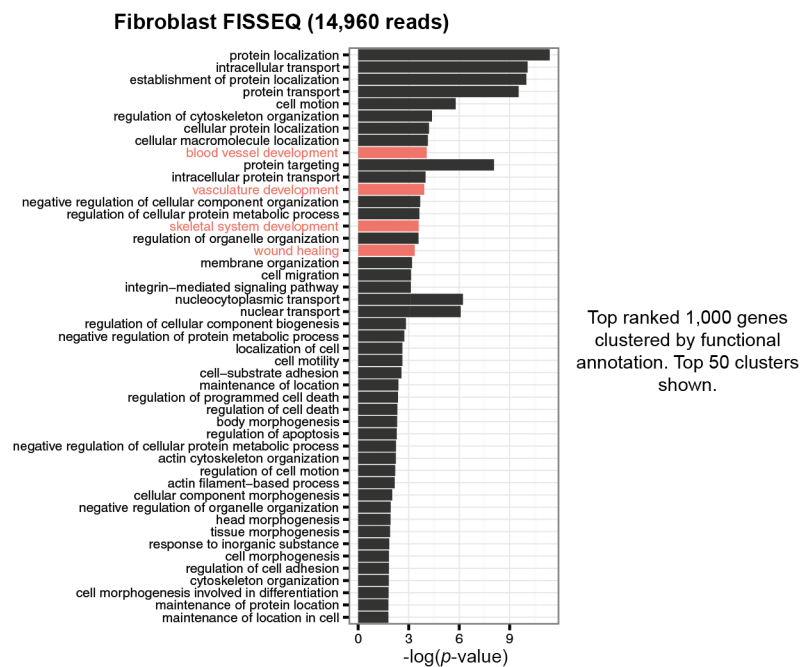
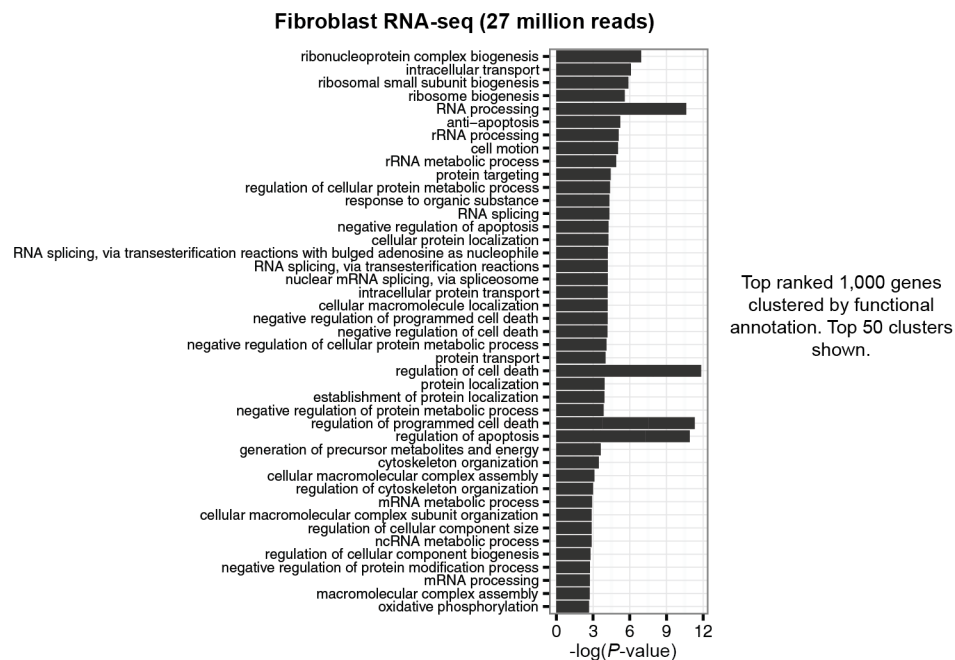


Figure S12

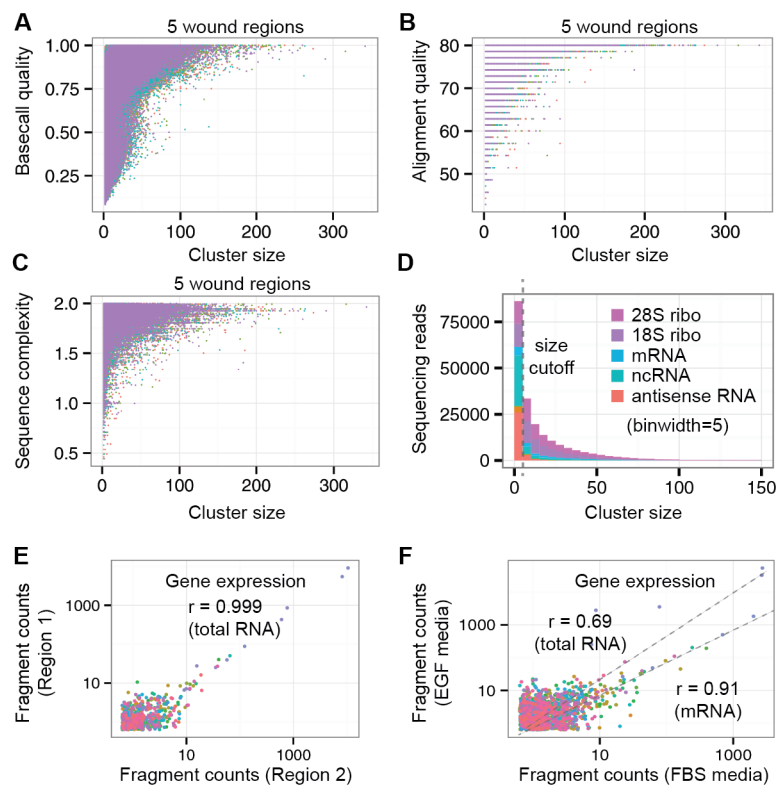


Figure S13

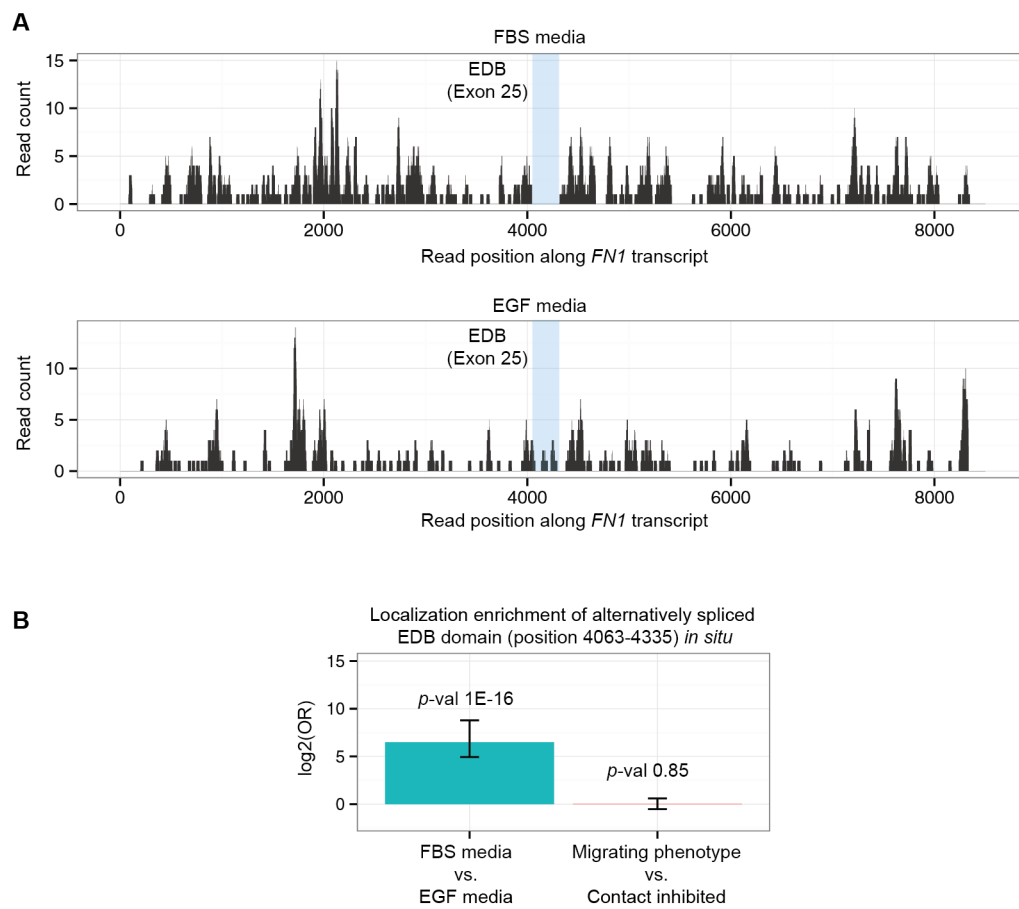


Figure S14

Table S1. FISSEQ summary statistics from human primary fibroblasts in FBS media.

Total bases from aligned amplicons (>5 pixels)	420,714 nt
Amplicons with >5 pixels	15,582 (100%)
Belongs to an annotated RNA class	15,126 (97.1%)
Maps to a single gene	14,960 (96.0%)
Unique gene names	4,171
Mean per-base error %	1.74%
Median per-base error %	0.64%

Table S2. The RNA localization likelihood compared to 16S mitochondrial rRNA.

Amplicons, 12,427 (83.1%) cytoplasmic and 2,533 (16.9%) nuclear, are compared against 16S rRNA localization. 164 genes with more than five observations are chosen for Fisher's exact test. All genes were more likely to be found in the nucleus compared to 16S mitochondrial rRNA, given their nuclear origin (odds ratio < 1). But non-coding RNA (*MALAT1*, *NEAT*, *KCNQ1OT1*), small nuclear RNA (*RN7SK*, *RNU2-1*), and pre-ribosomal RNA (*RNA45S5*) were notably more enriched in the nucleus.

Gene	Cytoplasmic	Nuclear	P-value	OR	95% CImin	95% CImax
MALAT1	295	164	1.2E-64	0.10	0.08	0.12
RIBO28S	1774	319	5.7E-29	0.31	0.25	0.37
RIBO18S	1493	258	9.0E-25	0.32	0.26	0.39
NEAT1	51	33	3.1E-19	0.09	0.05	0.12
RNA45S5	193	48	2.1E-13	0.23	0.15	0.30
RN7SK	75	19	1.0E-06	0.22	0.13	0.31
RNU2-1	18	9	8.3E-06	0.11	0.05	0.18
KCNQ1OT1	19	8	6.8E-05	0.13	0.05	0.21
GLS	7	5	2.7E-04	0.08	0.02	0.14
MYO10	5	4	8.5E-04	0.07	0.01	0.13
THBS1	186	23	1.7E-03	0.45	0.28	0.62
FNDC1	3	3	2.7E-03	0.06	0.01	0.10
MAP3K4	3	3	2.7E-03	0.06	0.01	0.10
VIM	22	6	3.4E-03	0.21	0.08	0.33
SIPA1L1	4	3	4.6E-03	0.07	0.01	0.14
VPS13C	4	3	4.6E-03	0.07	0.01	0.14
HNRNPA2B1	5	3	7.1E-03	0.09	0.02	0.17
SNORD50A	5	3	7.1E-03	0.09	0.02	0.17
DST	12	4	9.1E-03	0.17	0.05	0.29
CHD1	7	3	1.4E-02	0.13	0.03	0.23
IGFBP4	7	3	1.4E-02	0.13	0.03	0.23

Table S3. The RNA localization likelihood compared to *MALAT1*, a non-coding RNA known to localize to the nuclear speckles. Amplicons, 12,427 (83.1%) cytoplasmic and 2,533 (16.9%) nuclear, are compared against *MALAT1* localization. Also, 164 genes with more than five observations are chosen for Fisher's exact test. Most genes were more likely to be found in the cytoplasm compared to *MALAT1* (odds ratio > 1).

Gene	Cytoplasmic	Nuclear	<i>P</i> -value	OR	95% Cimin	95% Cimax
RIBO16S	2271	127	1.2E-64	9.93	7.58	12.28
RIBO12S	844	66	3.5E-38	7.10	5.14	9.06
FN1	481	29	1.9E-33	9.20	5.99	12.41
RIBO18S	1493	258	6.3E-22	3.22	2.53	3.90
RIBO28S	1774	319	2.1E-21	3.09	2.45	3.73
COL1A2	138	6	3.5E-16	12.75	5.53	19.97
COL1A1	160	14	1.4E-13	6.34	3.52	9.15
THBS1	186	23	5.4E-12	4.49	2.77	6.21
COL6A3	50	4	5.6E-06	6.93	2.48	11.38
COL6A1	29	0	6.4E-06	Inf	4.02	Inf
RNA45S5	193	48	1.4E-05	2.23	1.53	2.94
TGFB1	42	3	2.0E-05	7.76	2.42	13.11
SPARC	30	1	4.6E-05	16.63	2.71	30.54
ITGB1	38	3	9.2E-05	7.02	2.18	11.87
COL12A1	26	1	2.3E-04	14.41	2.32	26.50
COL3A1	28	2	5.4E-04	7.76	1.92	13.61
RNY3	18	0	5.9E-04	Inf	2.40	Inf
FBN1	35	4	7.0E-04	4.85	1.69	8.01
EEF1A1	30	3	1.0E-03	5.55	1.68	9.41

Table S4. The likelihood table of differentially expressed genes (180 genes with >5 observations) reveals biological pathway enrichment in migrating vs. stationary fibroblasts. For the purpose of generating fold-change plots (Fig. 4F), a small positive value (0.01) was added to normalized mRNA counts. Fisher's exact test used 28S rRNA counts in migrating cells vs. contact inhibited cells for comparison.

Gene	Migrating	Stationary	<i>P</i> -value	Pubmed ID
GID4	35	0	5.41E-10	PMID17467196, PMID18215130
CD300E	1	15	4.48E-03	PMC3009906
FHDC1	1	15	4.48E-03	PMID11779461
COL3A1	7	26	8.92E-03	
SSR1	12	2	1.79E-02	PMID22314232
LMO7	6	0	2.44E-02	PMC3147800
PRPF40A	6	0	2.44E-02	PMC3201212
WNK1	6	0	2.44E-02	PMID10660600
EEF2	0	8	2.58E-02	
ASPM	7	1	3.89E-02	PMID15972725
SLIT3	7	1	3.89E-02	PMID9813312
COL1A2	47	28	4.57E-02	

Movie S1. 3D rendering of the whole transcriptome amplicons in human iPS cells visualized using scanning confocal microscopy. Regions of low RNA abundance are observed as fluorescence-free region such as inter-cellular junctions and the nuclear membrane. In certain phases of the iPS cell growth, RNA in the nucleolus becomes inaccessible to our method, visible as a hollow dark center in the nucleus.

Movie S2. 3D rendering of the whole transcriptome amplicons and the early chromosome condensation in human iPS cells visualized using confocal microscopy.

Movie S3. 3D rendering of the whole transcriptome amplicons in a nucleolus-like body visualized using confocal microscopy. Typically, one nucleolus consists of ~100 cDNA amplicons in iPS cells.

Movie S4. Time series of 30 sequencing cycles in primary fibroblasts.

Movie S5. Time series of 30 sequencing cycles in primary fibroblasts (zoomed).

Movie S6. 3D rendering of the 30-base sequencing run in human primary fibroblasts. Time (sequencing base) is represented on the z-axis, allowing one to visualize each sequence as a vertical string of sequences within the cell.

References

1. G. Diez-Roux, S. Banfi, M. Sultan, L. Geffers, S. Anand, D. Rozado, A. Magen, E. Canidio, M. Pagani, I. Peluso, N. Lin-Marq, M. Koch, M. Bilio, I. Cantiello, R. Verde, C. De Masi, S. A. Bianchi, J. Cicchini, E. Perroud, S. Mehmeti, E. Dagand, S. Schrinner, A. Nürnberger, K. Schmidt, K. Metz, C. Zwingmann, N. Brieske, C. Springer, A. M. Hernandez, S. Herzog, F. Grabbe, C. Sieverding, B. Fischer, K. Schrader, M. Brockmeyer, S. Dettmer, C. Helbig, V. Alunni, M. A. Battaini, C. Mura, C. N. Henriksen, R. Garcia-Lopez, D. Echevarria, E. Puelles, E. Garcia-Calero, S. Kruse, M. Uhr, C. Kauck, G. Feng, N. Milyaev, C. K. Ong, L. Kumar, M. Lam, C. A. Semple, A. Gyenesei, S. Mundlos, U. Radelof, H. Lehrach, P. Sarmientos, A. Reymond, D. R. Davidson, P. Dollé, S. E. Antonarakis, M. L. Yaspo, S. Martinez, R. A. Baldock, G. Eichele, A. Ballabio, A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9**, e1000582 (2011). [Medline doi:10.1371/journal.pbio.1000582](#)
2. H. Zeng, E. H. Shen, J. G. Hohmann, S. W. Oh, A. Bernard, J. J. Royall, K. J. Glattfelder, S. M. Sunkin, J. A. Morris, A. L. Guillozet-Bongaarts, K. A. Smith, A. J. Ebbert, B. Swanson, L. Kuan, D. T. Page, C. C. Overly, E. S. Lein, M. J. Hawrylycz, P. R. Hof, T. M. Hyde, J. E. Kleinman, A. R. Jones, Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149**, 483–496 (2012). [Medline doi:10.1016/j.cell.2012.02.052](#)
3. E. Lécuyer, H. Yoshida, N. Parthasarathy, C. Alm, T. Babak, T. Cerovina, T. R. Hughes, P. Tomancak, H. M. Krause, Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174–187 (2007). [Medline doi:10.1016/j.cell.2007.08.003](#)
4. J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, G. M. Church, Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005). [Medline doi:10.1126/science.1117389](#)
5. J. B. Kim, G. J. Porreca, L. Song, S. C. Greenway, J. M. Gorham, G. M. Church, C. E. Seidman, J. G. Seidman, Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* **316**, 1481–1484 (2007). [Medline doi:10.1126/science.1137325](#)
6. R. D. Mitra, V. L. Butty, J. Shendure, B. R. Williams, D. E. Housman, G. M. Church, Digital genotyping and haplotyping with polymerase colonies. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5926–5931 (2003). [Medline doi:10.1073/pnas.0936399100](#)
7. R. Drmanac, A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, B. G. Kermani, P. Carnevali, I. Nazarenko, G. B. Nilsen, G. Yeung, F. Dahl, A. Fernandez, B. Staker, K. P. Pant, J. Baccash, A. P. Borcharding, A. Brownley, R. Cedeno, L. Chen, D. Chernikoff, A. Cheung, R. Chirita, B. Curson, J. C. Ebert, C. R. Hacker, R. Hartlage, B. Hauser, S. Huang, Y. Jiang, V. Karpinchyk, M. Koenig, C. Kong, T. Landers, C. Le, J. Liu, C. E. McBride, M. Morenzoni, R. E. Morey, K. Mutch, H. Perazich, K. Perry, B. A. Peters, J. Peterson, C. L. Pethiyagoda, K. Pothuraju, C. Richter, A. M. Rosenbaum, S. Roy, J. Shafto, U. Sharanhovich, K. W. Shannon, C. G. Sheppy, M. Sun, J. V. Thakuria, A. Tran, D. Vu, A. W. Zaranek,

- X. Wu, S. Drmanac, A. R. Oliphant, W. C. Banyai, B. Martin, D. G. Ballinger, G. M. Church, C. A. Reid, Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010). [Medline doi:10.1126/science.1181498](#)
8. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998). [Medline doi:10.1126/science.280.5363.585](#)
 9. J. M. Levsky, S. M. Shenoy, R. C. Pezo, R. H. Singer, Single-cell gene expression profiling. *Science* **297**, 836–840 (2002). [Medline doi:10.1126/science.1072241](#)
 10. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008). [Medline doi:10.1038/nmeth.1253](#)
 11. S. Itzkovitz, A. van Oudenaarden, Validating transcripts with probes and imaging technology. *Nat. Methods* **8** (Suppl), S12–S19 (2011). [Medline doi:10.1038/nmeth.1573](#)
 12. M. A. Frohman, M. K. Dush, G. R. Martin, Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998–9002 (1988). [Medline doi:10.1073/pnas.85.23.8998](#)
 13. R. J. McAnulty, Fibroblasts and myofibroblasts: Their source, function and role in disease. *Int. J. Biochem. Cell Biol.* **39**, 666–671 (2007). [Medline doi:10.1016/j.biocel.2006.11.005](#)
 14. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009). [Medline doi:10.1038/nprot.2008.211](#)
 15. A. R. Kornblihtt, C. G. Pesce, C. R. Alonso, P. Cramer, A. Srebrow, S. Werbajh, A. F. Muro, The fibronectin gene as a model for splicing and transcription studies. *FASEB J.* **10**, 248–257 (1996). [Medline](#)
 16. R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, M. Nilsson, In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013). [Medline doi:10.1038/nmeth.2563](#)
 17. K. Zhang, J. B. Li, Y. Gao, D. Egli, B. Xie, J. Deng, Z. Li, J. H. Lee, J. Aach, E. M. Leproust, K. Eggan, G. M. Church, Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat. Methods* **6**, 613–618 (2009). [Medline doi:10.1038/nmeth.1357](#)
 18. J. H. Lee, I. H. Park, Y. Gao, J. B. Li, Z. Li, G. Q. Daley, K. Zhang, G. M. Church, A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet.* **5**, e1000718 (2009). [Medline doi:10.1371/journal.pgen.1000718](#)
 19. J. Livet, T. A. Weissman, H. Kang, R. W. Draft, J. Lu, R. A. Bennis, J. R. Sanes, J. W. Lichtman, Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007). [Medline doi:10.1038/nature06293](#)