Highly Multiplexed Subcellular RNA Sequencing in Situ

Science
AAAS

This copy is for your personal, non-commercial use only.

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 Asp[138] 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 expressing 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 κB (NF-κB) 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-κB α (IκBα), 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. Moriwaki, 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).

Highly Multiplexed Subcellular RNA Sequencing in Situ

Je Hyuk Lee,[1,2]*† Evan R. Daugharthy,[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 requires 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

[1]Wyss Institute, Harvard Medical School, Boston, MA 02115, USA. [2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [3]Department of Stem Cell and Regenerative Biology, Harvard University, Boston, MA 02138, USA. [4]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. [5]Allen 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 (FN1); collagens (COL1A1, COL1A2, COL3A1); matrix metallopeptidases and inhibitors (MMP14, MMP2, TIMP1); osteonectin (SPARC); stanniocalcin (STC1); and the bone morphogenesis–associated transforming growth factor (TGF)–induced protein (TGFBI), 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 (FN1). 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).
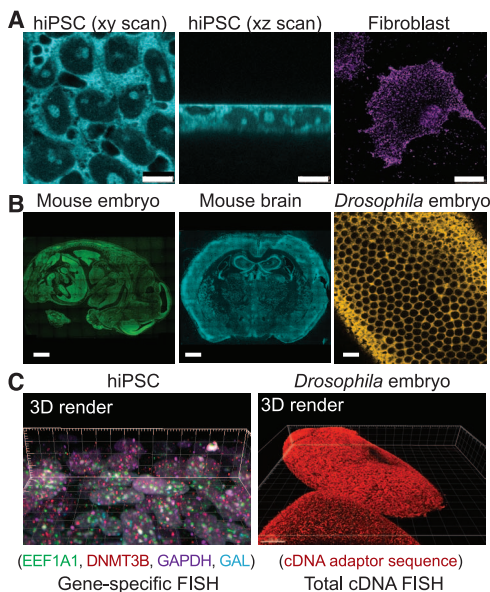

Fig. 1. Construction of 3D RNA-seq libraries in situ. After RT using random hexamers with an adapter sequence in fixed cells, the cDNA is amplified and cross-linked in situ. (**A**) A fluorescent probe is hybridized to the adapter sequence and imaged by confocal microscopy in human iPS cells (hiPSC) (scale bar: 10 μm) and fibroblasts (scale bar: 25 μm). (**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 μm), 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.
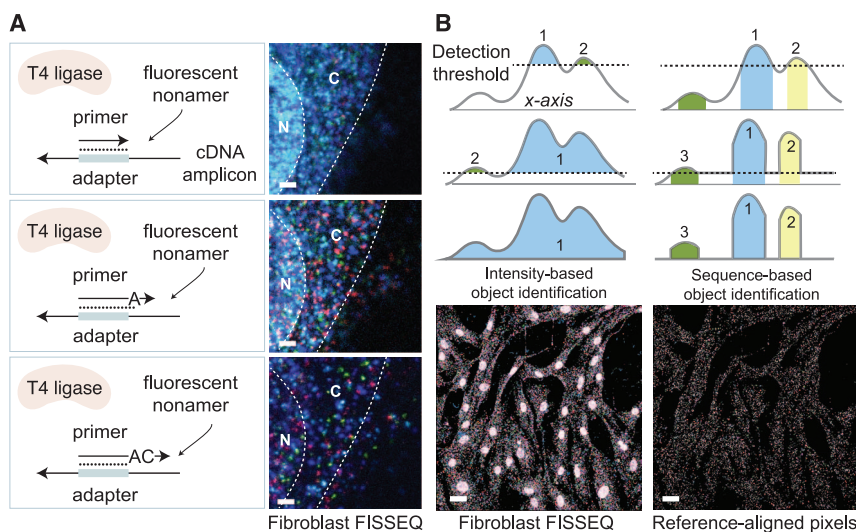

Fig. 2. Overcoming resolution limitations and enhancing the signal-to-noise ratio. (**A**) Ligation of fluorescent oligonucleotides occurs when the sequencing primer ends are perfectly complementary to the template. Extending sequencing primers by one or more bases, one can randomly sample amplicons at 1/4th, 1/16th, and 1/256th of the original density in fibroblasts (scale bar: 5 μm). 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 μm).

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 noncoding ($P < 10^{-16}$), 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).
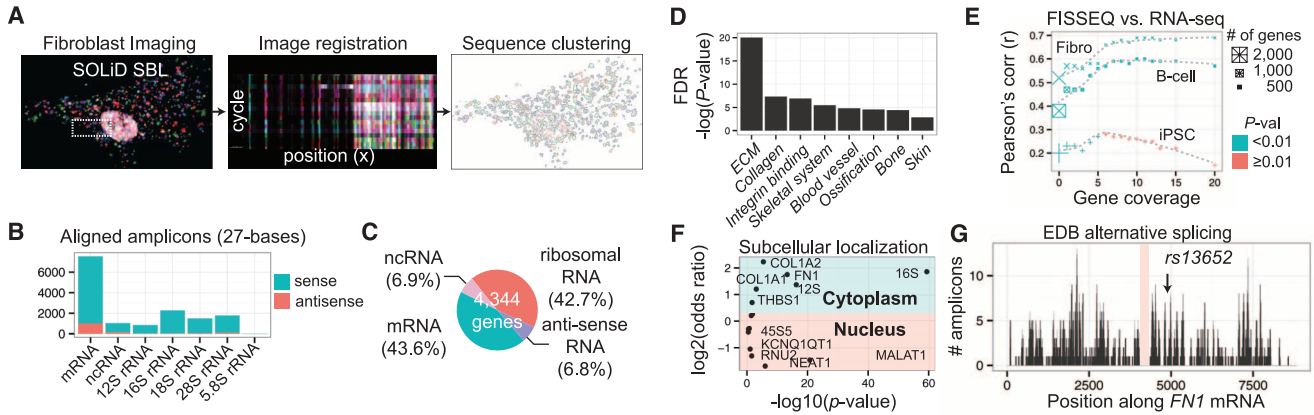

Fig. 3. Whole-transcriptome in situ RNA-seq in primary fibroblasts. (**A**) From deconvolved confocal images, 27-base reads are aligned to the reference, and alignments are spatially clustered into objects. (**B**) Of the amplicons, 90.6% align to the annotated (+) strand. (**C**) mRNA and non-coding 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, thrombospodin; COMP, cartilage oligomeric matrix protein; CHAD, chondroadherin; 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.
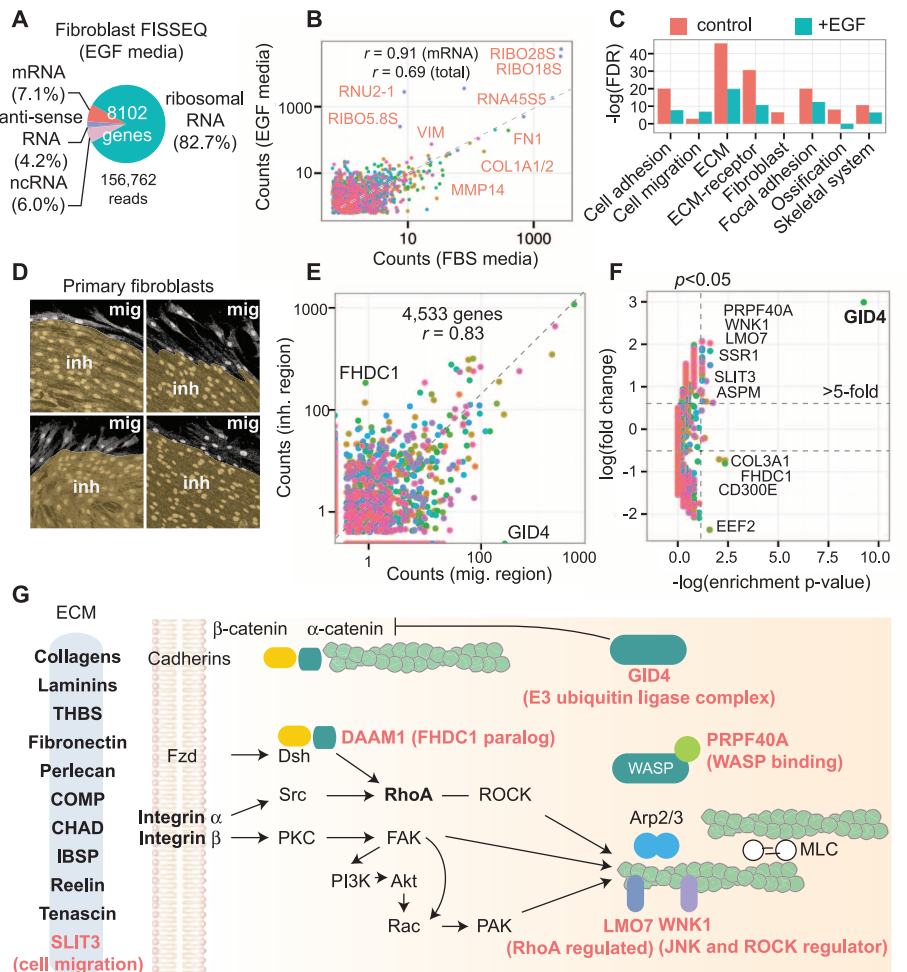
FN1 has three variable domains referred to as EDA, EDB, and IIICS, which are alternatively spliced (*15*). We did not observe development-associated EDB, but observed adult tissue–associated EDA and IIICS (Fig. 3G).

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 *FN1* (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. Levsky, 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. Kornblihtt *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).

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 cholesterol 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 mitochondrial 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,

[1]Max-Planck-Institut für Biophysikalische Chemie, 37077 Göttingen, Germany. [2]Deutsches Zentrum für Neurodegenerative Erkrankungen, 37077 Göttingen, Germany. [3]Center 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.mpibpc.mpg.de (S.B.)