GENETICS

NanoSINC-seq dissects the isoform diversity in subcellular compartments of single cells

Yusuke Oguchi[1,2], Yuka Ozaki[1], Mahmoud N. Abdelmoez[1], Hirofumi Shintaku[1]*

Alternative mRNA isoforms play a key role in generating diverse protein isoforms. To dissect isoform usage in the subcellular compartments of single cells, we introduced an novel approach, nanopore sequencing coupled with single-cell integrated nuclear and cytoplasmic RNA sequencing, that couples microfluidic fractionation, which separates cytoplasmic RNA from nuclear RNA, with full-length complementary DNA (cDNA) sequencing using a nanopore sequencer. Leveraging full-length cDNA reads, we found that the nuclear transcripts are notably more diverse than cytoplasmic transcripts. Our findings also indicated that transcriptional noise emanating from the nucleus is regulated across the nuclear membrane and then either attenuated or amplified in the cytoplasm depending on the function involved. Overall, our results provide the landscape that shows how the transcriptional noise arising from the nucleus propagates to the cytoplasm.

INTRODUCTION

Protein-coding transcripts are synthesized and processed in the nucleus and exported to the cytoplasm where translation occurs. Alternative isoform usage in the cytoplasm is essential for the diversification of protein molecules, even in isogenic cell populations (*1*, *2*). Furthermore, the isoforms modulate the functions of long noncoding RNAs (lncRNAs) enriched in the nucleus via the inclusion or exclusion of specific domains (*3*). Our current understanding on nucleocytoplasmic isoform usage, combined with nuclear export and degradation of transcripts, is limited at the population-level resolution (*4*–*7*), and its effects on cellular heterogeneity at the single-cell resolution remain largely unexplored. Thus, to better understand isoform localization and usage at single-cell resolution, we leveraged physical fractionation, which separates cytoplasmic RNAs from nuclear RNA, via a microfluidic approach coupled with Oxford Nanopore Technologies (ONT) long-read sequencing. We termed this approach nanopore sequencing coupled with single-cell integrated nuclear and cytoplasmic RNA sequencing (RNA-seq) (NanoSINC-seq). Our findings showed that nuclear transcript usage is more diverse than the cytoplasmic transcript usage in leukemic cells (K562), providing sources to trigger fate commitment via transcriptional bursting. Diversity of transcript usage is fine-tuned to attenuate or amplify the transcriptional noise in the cytoplasm being coordinated toward specific cellular functions.

RESULTS

NanoSINC-seq workflow

We prepared cytoplasmic RNA and nuclear RNA from the same single K562 cells using the microfluidic SINC-seq system (Fig. 1). We have previously described the working principle underlying the microfluidic system and the procedures used to critically benchmark it (*8*–*11*). Briefly, this approach leverages the electric field in a microfluidic channel to selectively lyse the plasma membrane of a single cell, while keeping the nucleus intact, resulting in stringent fractionation into cytoplasmic RNA and nuclear RNA via electrophoretic separation (*12*, *13*). Fractionated RNA molecules were extracted from the microfluidic system using a standard micropipette, following which the SMART-Seq v4 protocol was applied to synthesize complementary DNA (cDNA) products from polyadenylated RNA molecules in individual polymerase chain reaction (PCR) tubes.

To explore the isoform diversity in the cytoplasm and nuclei of single cells, we constructed Illumina and ONT sequencing libraries using identical cDNA products (Fig. 1). In total, 16 pairs of cytoplasmic and nuclear RNA-seq libraries were prepared from 16 single cells and 8 single-cell libraries for both Illumina and ONT sequencing. We sequenced the barcoded ONT libraries with eight flow cells by pooling up to five samples (two of cytoplasmic RNA, two of nuclear RNA, and one single cell) per run. The sequencing run yielded 2.2 ± 1.4 M reads per pooled sample (fig. S1A), of which 62.5 ± 11% was successfully demultiplexed with a pair of 30–base pair (bp)–long barcodes (table S1). We discarded the undemultiplexed reads and did not use them in subsequent analyses. Here, we downsampled demultiplexed reads of cytoplasmic or nuclear cDNA to compare them at a similar depth (fig. S1B).

Benchmarking NanoSINC-seq in comparison to SINC-seq

NanoSINC-seq detected 4538 ± 1950 and 4932 ± 1907 transcripts in the cytoplasmic and nuclear fractions of single cells, respectively. Among which, 2011 ± 940 transcripts were common, corresponding to 3642 ± 1297 and 3915 ± 1216 genes from the respective fractions, of which 2114 ± 911 were common (Fig. 2, A and B). The marker genes of subcellular compartments were enriched in cytoplasmic and nuclear marker genes in the respective fractions (fig. S1, C to E, and table S2). Gene expression detected via NanoSINC-seq and SINC-seq based on the Illumina platform were quantified and compared. Gene expression in the cytoplasm, nuclei, and single cells was consistent across different platforms with relatively high correlation at the gene level and moderate correlation at the transcript level (fig. S1, F and G). We evaluated the effect of sequence depth on the detection of transcripts and genes via the NanoSINC-seq (fig. S1, H and I) and compared them with those of bulk RNA samples in the Supplementary Materials (fig. S2, A and B). Under the same depths, NanoSINC-seq showed comparable performance in detecting similar numbers of transcripts (fig. S1F) and genes (fig. S1G) in the two fractions; notably, more transcripts and genes were

[1]Cluster for Pioneering Research, RIKEN, 2-1 Hirosawa, Wako, Saitama, Japan. [2]PRESTO, Japan Science and Technology Agency, 4-1-8, Honcho, Kawaguchi, Saitama 332-0012, Japan.
*Corresponding author. Email: hirofumi.shintaku@riken.jp

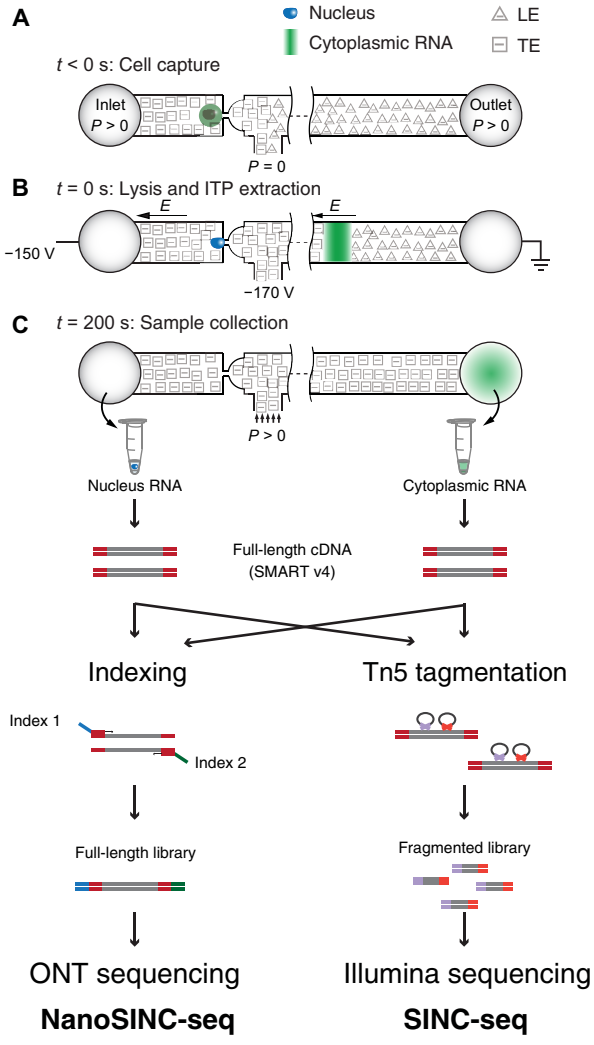

Fig. 1. NanoSINC-seq. Workflow of NanoSINC-seq. (**A**) Single-cell isolation at a hydrodynamic trap via pressure-driven flow ($t < 0$ s). (**B**) Lysis of cell membrane and cytoplasmic RNA extraction via electrophoretic nucleic acid extraction ($t = 0$ s). (**C**) Sample extraction and library preparation for sequencing with long (ONT) and short (Illumina) read sequencers. Leading electrolyte (LE) and trailing electrolyte (TE) are aqueous buffers for the electrophoretic extraction with isotachophoresis (ITP) (*12*, *13*, *32*). *P* and *E* represent the pressure and electric field, respectively.

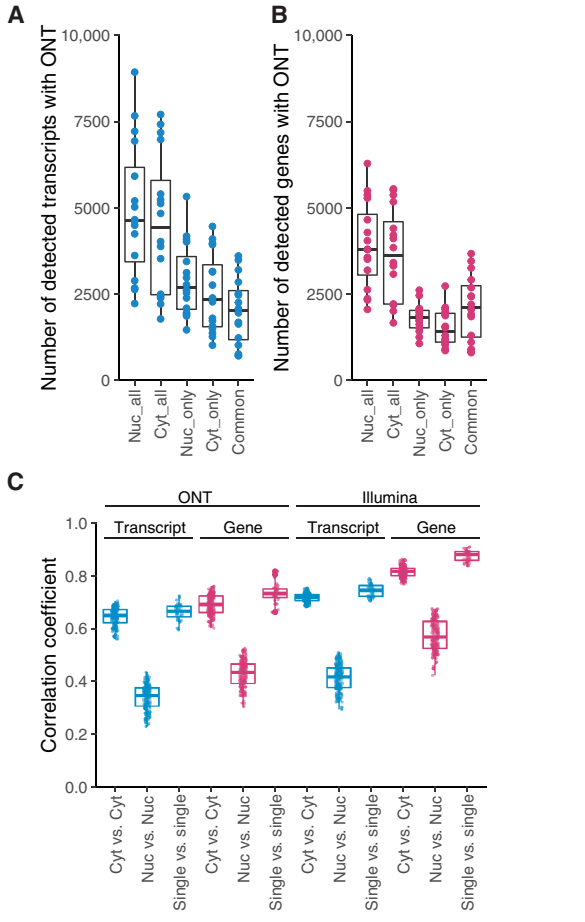

Fig. 2. Benchmark of NanoSINC-seq assessing detection of transcripts and comparing to SINC-seq. (**A**) Numbers of transcripts detected via ONT. (**B**) Number of genes detected via ONT. Cyt and Nuc indicate cytoplasmic and nuclear fractions, respectively. "_all" indicates all of the detected transcripts/genes in individual fractions. "_only" indicates transcripts/genes detected only in one of the fractions. "Common" indicates detected transcripts/genes both in nuclear and cytoplasmic fractions. (**C**) Differences in the coefficients of correlation at transcript level and gene level, obtained via ONT sequencing. Each coefficient of correlation was calculated for a pair of nuclear fractions, a pair of cytoplasmic fractions, and a pair of single cells, respectively. A similar analysis was performed for Illumina sequencing.

cumulatively detected in the nuclear fraction than the cytoplasmic fraction (fig. S2, C and D). This observation was consistent with lower numbers of overlapping transcripts and genes in pairs of replicated nuclear RNA-seq than those of cytoplasmic RNA-seq (fig. S2E). These suggest that the expression of transcript and gene in the nuclear fraction is more diverse than that in the cytoplasmic fraction.

We next assessed the effect of transcript length on quantification. In general, NanoSINC-seq showed no notable length bias compared with SINC-seq, except for transcripts shorter than 900 bp (fig. S3A). Byrne *et al.* (*14*) have indicated that ONT sequencing overrepresented the expression of short transcripts (<500 bp). In contrast, NanoSINC-seq underrepresented the expression of short transcripts (<900 bp). This discrepancy may be attributed to differences in the ONT sequencing library configuration, i.e., the two-dimensional (2D) protocol used by Byrne *et al.* (*14*) and the 1D protocol we used. (Note that the 2D protocol is currently unavailable from ONT.) To elucidate the difference between long- and short-read sequencing, we performed a differential analysis of ONT and Illumina sequencing at the transcript level using the respective sample types and a threshold of abs $\{\log_2[\text{fold change (FC)}]\} > 1$ and $P < 0.05$ (Wilcoxon signed-rank test, adjusted with a Benjamini-Hochberg correction; fig. S3, B and C). Transcripts enriched in ONT sequencing were longer and contained longer exons than those in Illumina sequencing regardless of sample type (fig. S3, D to F). This dependence on length may be due to systematic amplification or data analysis bias. However, NanoSINC-seq successfully captured unique cell-to-cell variabilities specific to three cellular compartments as evidenced by coefficients of correlation. We found that the lowest correlation among the three cellular compartments is at the nuclear level and that more moderate correlation within the same cellular compartment is at the transcriptional level in comparison to the gene level (Fig. 2C). Thus, observations derived from NanoSINC-seq

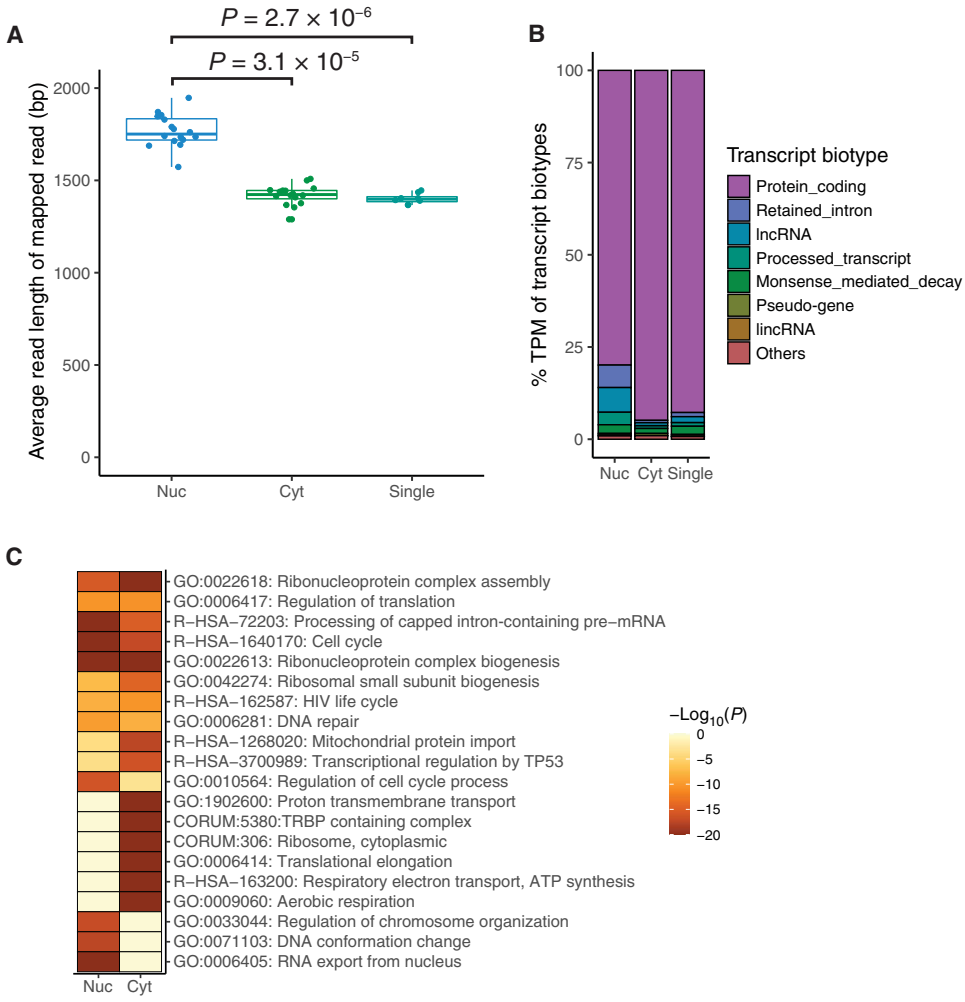

Fig. 3. NanoSINC-seq depicts the difference in transcripts between cytoplasmic and nuclear fractions. (**A**) Mean length of aligned reads of the three sample types: nuclei, cytoplasm, and single cells. The statistical test was performed with two-sided Wilcoxon signed-rank test for Nuc ($n = 16$) versus Cyt ($n = 16$) and two-sided single Wilcoxon rank sum test for Nuc ($n = 16$) versus single ($n = 8$). (**B**) Proportion of transcript biotypes of the three sample types: nuclei, cytoplasm, and single cells. The statistical analyses are provided in fig. S5. lincRNA, long intervening noncoding RNA. (**C**) Gene ontology (GO) analysis of DETs in the nuclei and cytoplasm by Metascape (*31*) . ATP, adenosine 5′-triphosphate. TPM, transcripts per million; TP53, tumor protein p53; TRBP, transactivation response RNA binding protein.

data were consistent with those based on SINC-seq. We provided further benchmarking of NanoSINC-seq in Supplementary Text and fig. S4.

Transcriptional differences between the cytoplasm and nucleus of single cells

To identify transcriptional differences across cellular compartments, we leveraged NanoSINC-seq data to explore the difference between cytoplasmic and nuclear transcripts. The mean length of the aligned reads from nuclear transcripts was longer than that from cytoplasmic transcripts or single cells (Fig. 3A). This is partially explained by the enriched intronic regions of nascent transcripts in the nucleus (*15*). Transcript biotypes in nuclear transcripts clearly showed distinct compositions, increased expression of retained introns, lncRNA, and processed transcripts in comparison with those in cytoplasmic and whole-cell (single-cell) transcripts (Fig. 3B and fig. S5). Next, the differences between the expression levels of cytoplasmic and nuclear transcripts were explored (fig. S6A). Of the 29,256 transcripts cumulatively detected in the 16 pairs of samples, we identified 849 and 574 differentially enriched transcripts (DETs) in the cytoplasmic and nuclear compartments, respectively, using a threshold of abs [$\log_2$(FC)] > 1 and $P < 0.05$ (Wilcoxon signed-rank test adjusted with a Benjamini-Hochberg correction). Gene ontology analysis revealed that compared with those in the cytoplasm, transcripts enriched in the nuclei are associated with DNA conformation changes, RNA export from the nucleus, and the regulation of chromosome organization (Fig. 3C). Meanwhile, transcripts enriched in the cytoplasm are associated with ribosomes and translational elongation. The biogenesis of the ribonucleoprotein complex was enriched in both fractions, reflecting the assembly and maturation processes of ribosomes, which involve nucleocytoplasmic transport (*16*). These findings from NanoSINC-seq were consistent with those from SINC-seq (fig. S6B). Among the genes localized in the subcellular compartments, genes expressing multiple DETs were enriched in specific compartments (fig. S6, C to E).

We additionally performed a statistical comparison on the number of isoforms detected per gene between the cytoplasm and nucleus [Wilcoxon signed-rank test adjusted with a Benjamini-Hochberg

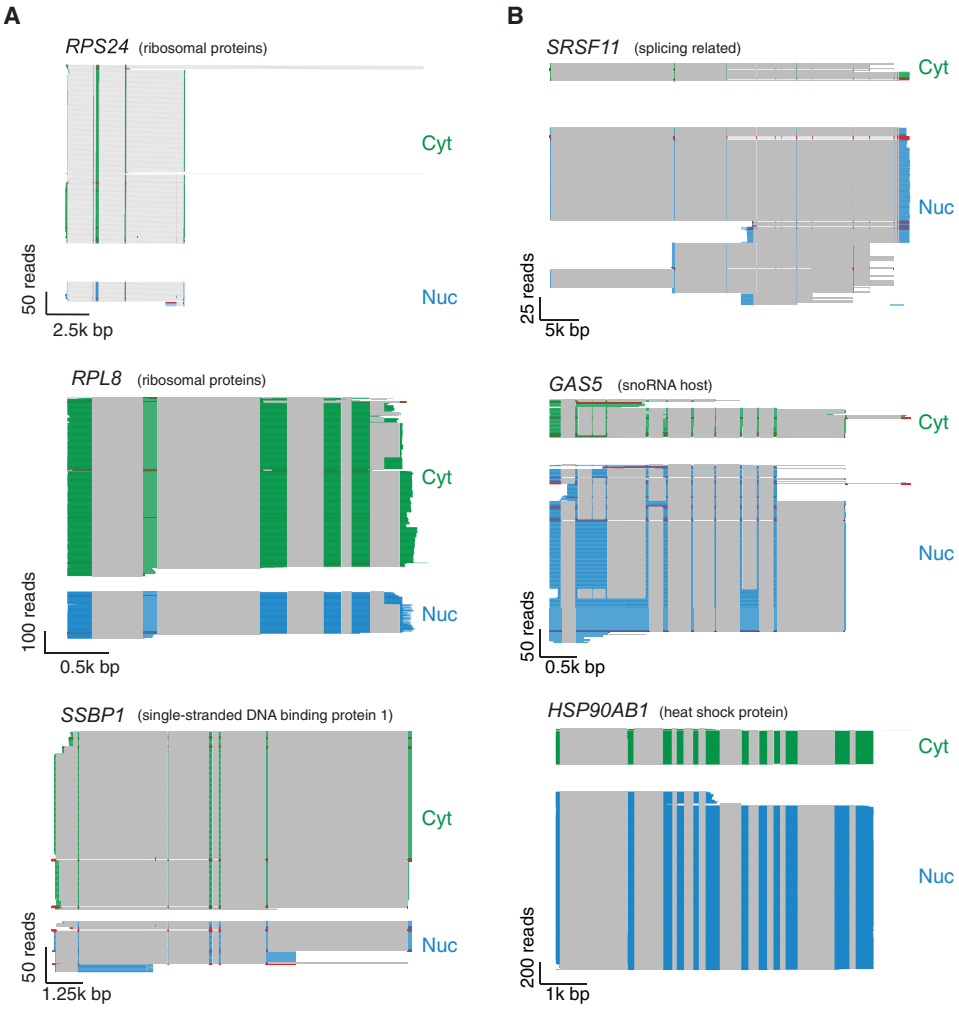

Fig. 4. Genes showing high diversity in the isoform usage in either cytoplasmic or nuclear fraction. Transcript coverage in the cytoplasmic (**A**) and nuclear DETs (**B**). Genes expressing multiple isoforms that were enriched in cytoplasm and nuclei (fig. S6C). Statistical analyses on the detection of multiple isoforms and enrichment are described in fig. S7. The isoforms were identified using a computational workflow that identifies transcripts with high confidence (FLAIR) (*30*).

correction ($P < 0.05$); fig. S7] and identified 296 and 293 isoforms with higher diversity in the isoform usage in cytoplasmic and nuclear fractions, respectively. For instance, ribosomal protein–related genes (*RPS24* and *RPL8*) and DNA binding protein gene (*SSBP1*) were enriched in the cytoplasm, whereas splicing (*SRSF11*), small nucleolar RNA (snoRNA) host (*GAS5*), and heat shock protein (*HSP90AB1*) were enriched in the nucleus (Fig. 4, A and B). Comparing the isoforms with no significant localization {abs[$\log_2$(FC)] < 1 or $P > 0.05$}, we observed slightly more diverse isoform usage in cytoplasmic fraction (fig. S7B).

Amplification and attenuation of transcriptional noise arising from nuclei

We explored cell-to-cell variability at transcriptional resolution using the NanoSINC-seq dataset. In eukaryotic cells, nuclear export modulates the transcriptional abundance in the cytoplasm by attenuating or amplifying transcriptional noise arising from transcriptional bursts in the nucleus. Battich *et al.* (*17*) and Bahar Halpern *et al.* (*18*) have proposed that the compartmentalization of transcripts through the nuclear membrane reduces cell-to-cell variation in the cytoplasm by acting as a passive noise filter. Conversely, Hansen *et al.* (*19*) have suggested that the nuclear export amplifies the variation in the cytoplasm compared with that in the nucleus. These studies quantified the abundance of transcripts in the cellular compartments using RNA fluorescence in situ hybridization; however, only a limited number of genes were targeted.

NanoSINC-seq data can illuminate the transcriptional fluctuation landscape in the two subcellular compartments. To compare the magnitude of cell-to-cell variation in the two compartments, we computed the squared variance to mean ratio (σ^2/μ) known as the Fano factor that effectively scales the decrease in coefficients of variation with decreasing means (fig. S8). NanoSINC-seq data revealed both amplified and attenuated cases at the transcriptional resolution (Fig. 5A). We compared the variation in gene resolution statistically by measuring the Fano factor at the gene level (Fig. 5B). This revealed 241 amplified and 104 attenuated genes in the cytoplasm due to nuclear compartmentalization. Notably, a comparison between the amplified and attenuated genes showed that the nuclear fraction has distinct compositions of biotypes (Fig. 5C and fig. S9). Gene ontology analyses revealed even further distinctions between

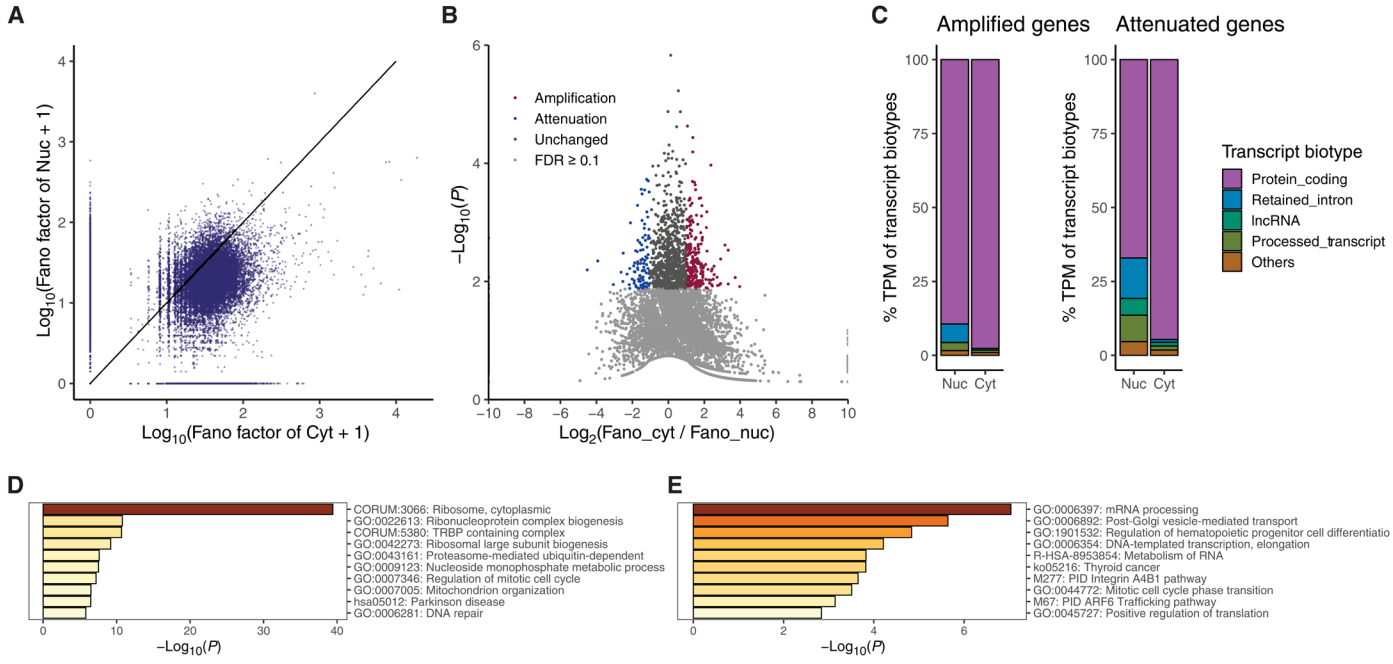

Fig. 5. Comparison of cell-to-cell variability between cytoplasmic transcripts and nuclear transcripts. (**A**) Fano factors corresponding to cytoplasmic transcripts are plotted against corresponding nuclear transcripts. The identity line separates attenuated (top) and amplified (bottom) transcripts. (**B**) Statistical comparison of variability at gene resolution level. Fano factors at the gene level are calculated by averaging Fano factors at the transcript level. Genes showing amplified variation in the cytoplasm are on the right-hand side. Amplified genes (red) were identified as those exhibiting *P* values less than 0.05 and $\log_2$(FC) greater than unity, and the attenuated genes (blue) were identified as those exhibiting *P* values less than 0.05 and $\log_2$(FC) less than −1. [*t* test adjusted with a Benjamini-Hochberg correction; false discovery rate (FDR) < 0.1; $n = 16$ each]. (**C**) Proportion of transcriptional biotypes of amplified genes and attenuated genes. (**D** and **E**) Gene ontology analysis of amplified genes and attenuated genes. The statistical test was performed using Metascape (*31*).

the amplified and attenuated genes (Fig. 5, D and E), indicating that transcriptional noise is specifically modulated (either attenuated or amplified) across the nuclear membrane.

DISCUSSION

NanoSINC-seq enables the exploration of isoform usage in the cytoplasm and nuclei of single cells by leveraging physical fractionation of subcellular RNAs and full-length cDNA sequencing without isoform identification based on a statistical model. Profiling of full-length cDNA successfully highlighted distinct isoform diversity and variance in transcript abundance in the subcellular compartments. The functions of localized transcripts enriched in the cytoplasm, and the nucleus were congruent with the results obtained via SINC-seq based on Illumina sequencing but were derived with a resolution at the transcript level.

NanoSINC-seq revealed that the transcriptional variability in the nucleus was higher than that in the cytoplasm (Fig. 2C). We hypothesized that the coefficient of variation reflects biological variability rather than the poor quantification obtained using the NanoSINC-seq pipeline based on four reasons. First, the correlation between ONT and Illumina sequencings were consistent among different types of single-cell samples (fig. S1E), suggesting insignificant bias regarding sample types. Second, our microfluidic approach can offer a similar sensitivity and reproducibility in detecting nuclear gene expression to a detergent-based approach using an off-the-shelf kit (PARIS Kit, Thermo Fisher Scientific) (*8*). Third, the amount of mRNA in the nucleus is relatively small, which accounts for ~16% of the whole mRNA in a single K562 cell (*8*). Fourth, the variability in single nuclear RNA-seq of neurons measured on a Fluidigm C1 platform using Illumina sequencing was reported as $r = 0.51$ to 0.62 (*20*). These findings suggest that the relatively high variability in nuclear gene expression was due to biological characteristics of nuclear RNA expression. In addition, NanoSINC-seq revealed that transcriptional and gene diversities in the nucleus were another source of cellular variability. NanoSINC-seq cumulatively detected more transcripts and genes in the nucleus than in the cytoplasm, despite the similarity in the number of transcripts detected per nucleus and per cytoplasm under similar read depth.

Moreover, NanoSINC-seq allowed the exploration of the landscape of posttranscriptional modulation to translation and addressed whether gene expression fluctuation across the nuclear membrane is amplified or attenuated by cellular processes. Cell-to-cell variability quantified with our data showed that 241 genes were amplified in the cytoplasm and 104 were attenuated. Notably, both amplified and attenuated genes displayed distinct biotype compositions and functions. Cumulatively, the amplification and attenuation of transcriptional fluctuations that arise in the nucleus propagate to the cytoplasm depending on biological function.

We sought for the overlap between the localization in a subcellular compartment and the amplified/attenuated genes and found small fractions of overlapping genes. Among the cytoplasmic enriched genes, 78 showed amplification, and none were attenuated. In contrast, among the nuclear-enriched genes, 11 were amplified, and 66 were attenuated. These findings imply a weak relationship between the localization and noise control; however, further studies

are needed to confirm the said relationship and identify its biological implications.

The proposed NanoSINC-seq approach has intrinsic limitations, particularly in terms of throughput and read depth mainly due to nanopore sequencing with MinION. This can be partly improved using higher-throughput instruments, such as GridION or PromethION from ONT. Alternatively, Smart-seq3 (*21*), which enables the in silico reconstruction of isoforms using a 5′ unique molecular identifier (UMI) RNA counting strategy and short read–based full-length coverage, can be readily integrated with our proposed microfluidic approach. We expected that this would improve the limitations due to sequencing depth of long-read approaches. Furthermore, we are currently developing a new microfluidic system that parallelly processes single cells to fractionate cytoplasmic RNA and nuclear RNA. We hope to address the limitations of this study in our ongoing project. In addition, the use of UMI with our ONT sequencing protocol poses another challenge due to relatively high error rate. Advances in the implementation of UMI (*22*, *23*) will improve the precision of quantitation and allow more accurate quantification of transcriptional noise. The NanoSINC-seq approach is readily applicable to various types of mammalian cells with a slight optimization of the applied voltage. We thus envision that it helps to better understand the isoform localization in disease cells, for example, brain disorders applying NanoSINC-seq to neural cells.

MATERIALS AND METHODS

Microfluidic fractionation

The microfluidic protocol enabled subcellular fractionation by coupling electrical lysis and isotachophoresis (ITP)–based extraction of nucleic acids. Cytoplasmic RNA and nucleus were extracted in different reservoirs, and the entire process was completed in less than 5 min (Fig. 1A). The T-shaped microchannel used was developed in our previous work (*8*). Briefly, we injected ITP buffers, leading electrolyte (LE) and trailing electrolyte (TE), into the channel to form an LE-TE interface at the T-junction. The LE buffer contained 25 mM HCl and 50 mM tris (Sigma-Aldrich), and the TE buffer contained 25 mM Hepes (Dojindo Laboratories, Kumamoto, Japan) and 50 mM imidazole (Sigma-Aldrich). To suppress electroosmotic flow, 0.4% polyvinylpyrrolidone (K90; PanReac AppliChem, Darmstadt, Germany) was added to both buffers. K562 human myelogenous leukemia cells were maintained in F-12 Nutrient Mix (Gibco, Grand Island, NY, USA) supplemented with 0.1× penicillin-streptomycin (Sigma-Aldrich), 10% fetal bovine serum at 37°C, and 5% CO_2. Before loading the cells into the channel, they were suspended in TE buffer supplemented with 175 mM sucrose (Wako Pure Chemical Industries). Cell lysis and RNA fractionation were achieved by voltage control with end-channel electrodes through custom-made MATLAB scripts. This procedure has been described previously (*8*), and a similar protocol with a narrated video is available (*24*). We note that, while we focus the fractionation of cytoplasmic RNA versus nuclear RNA in this study, our approach may be extended to the fractionation of subcellular proteins with some modifications (*25*, *26*).

Synthesis of cDNA from the fractionated nucleus and cytoplasmic RNA

We used the SMART-seq v4 Ultra Low Input RNA Kit (Takara Bio, USA) to synthesize cDNA from the fractionated RNAs in individual PCR tubes according to the manufacturer's protocol. We performed 18 cycles of PCR amplification.

Illumina sequencing and data analysis

The Illumina library was prepared using the identical cDNA product with the Nextera XT DNA Sample Preparation Kit (Illumina, USA). The library was sequenced on HiSeq2500 (Illumina, USA) with 100-base paired-end reads and an average depth of 2.4 ± 0.36 M reads. We mapped the sequencing reads to the human reference genome (GRCh37.75) using the STAR (v2.5.1b) mapping program (*27*) with ENCODE options and estimated the expression with transcripts per million using RNA-seq via Expected-Maximization (RSEM v1.3.0) (*28*). RSEM outputs expression at both the gene and isoform levels with a gtf file (Homo_sapiens.GRCh37.75.gtf), which was identical with full-length alternative isoform analysis of RNA (FLAIR). The precision of isoform identification and quantification using single-cell RNA-seq data by RSEM has been previously validated by Westoby *et al.* (*29*). We thus used expression at the transcript level for Illumina sequencing with estimated_isoform_results obtained by RSEM, which was then compared with those obtained via ONT sequencing.

Nanopore sequencing and data analysis

We constructed an ONT sequence library by amplifying 0.5 ng of cDNA with KAPA HiFi HotStart (KK2601) and a pair of multiplexing primers (6 pmol each; table S1). We incubated 20 μl of the reaction mixture at 95°C for 3 min, followed by 13 cycles at 98°C for 20 s, 67°C for 15 s, and 72°C for 4 min with a final extension at 72°C for 5 min. We pooled up to five samples and purified them using ×1.8 AMPure beads. Next, sequencing adaptors were ligated with 0.2 pmol (~220 ng) of the cDNA amplicon using the Ligation Sequencing Kit 1D (SQK-LSK108) following the manufacturer's protocol. The library was then purified with ×1.8 AMPure beads. We ran a 48-hour sequencing protocol with the pooled library using a single flow cell (FLO-MIN106.1 R9.4) on a MinION. The obtained sequencing reads were demultiplexed as follows: First, we trimmed 100 bp from both ends of each read, which were expected to contain a barcode sequence; next, we mapped the trimmed region to the barcode ID reference (table S1) with BLAST (2.2.30+; option: blastn -db barcode.id.reference -query trimed.region.fastq -out out.file -word_size 4 -num_descriptions 1 -num_alignments 1 -dust no -strand plus -outfmt "7 std qlen qseq sseq") and identified barcode IDs with $e < 10^{-5}$. We adopted the reads identified with the correct pair of barcodes for further analyses.

We aligned the demultiplexed reads to a reference genome (GRCh37.75 and Homo_sapiens.GRCh37.75.gtf.) using GraphMap (v0.5.2) with the default option (graphmap align -r reference.fa --gtf reference.gtf -d reads.fastq -o out.sam). The obtained sam format files were converted into the bam format, downsampled (fig.S1B) with Samtools (v1.9), and further converted into the bed12 format with Bedtools (v2.26.0). We quantified the expression per transcript and gene using FLAIR (v1.4.0) (*30*) with flair.py correct, flair.py collapse, and flair.py quantify. We found that FLAIR failed to quantify reads even when correctly mapped to the important marker genes of cellular compartments such as mitochondrial RNA. We rescued those unannotated reads using an in-house script. Transcript information about exon lengths and transcript biotypes was extracted from Homo_sapiens.GRCh37.75.gtf. We performed the gene ontology analysis using Metascape (*31*).

We calculated the Fano factor as squared variance over the mean of expression at isoform levels. We then identified amplified genes,

which are highly variable in the cytoplasmic fraction, as those having more than twofold variance and $P < 0.05$ [t test adjusted with a Benjamini-Hochberg correction; false discovery rate (FDR) < 0.1] and attenuated genes, which are highly variable in the nuclear faction.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/15/eabe0317/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. A. Arzalluz-Luque, A. Conesa, Single-cell RNAseq for the study of isoforms—How is that possible? *Genome Biol.* **19**, 110 (2018).
2. K. Karlsson, S. Linnarsson, Single-cell mRNA isoform diversity in the mouse brain. *BMC Genomics* **18**, 126 (2017).
3. T. R. Mercer, D. J. Gerhardt, M. E. Dinger, J. Crawford, C. Trapnell, J. A. Jeddeloh, J. S. Mattick, J. L. Rinn, Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* **30**, 99–104 (2011).
4. S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
5. K. Ninomiya, N. Kataoka, M. Hagiwara, Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. *J. Cell Biol.* **195**, 27–40 (2011).
6. L. Chen, A global comparison between nuclear and cytosolic transcriptomes reveals differential compartmentalization of alternative transcript isoforms. *Nucleic Acids Res.* **38**, 1086–1097 (2009).
7. J. M. Taliaferro, M. Vidaki, R. Oliveira, S. Olson, L. Zhan, T. Saxena, E. T. Wang, B. R. Graveley, F. B. Gertler, M. S. Swanson, C. B. Burge, Distal alternative last exons localize mRNAs to neural projections. *Mol. Cell* **61**, 821–833 (2016).
8. M. N. Abdelmoez, K. Iida, Y. Oguchi, H. Nishikii, R. Yokokawa, H. Kotera, S. Uemura, J. G. Santiago, H. Shintaku, SINC-seq: Correlation of transient gene expressions between nucleus and cytoplasm reflects single-cell physiology. *Genome Biol.* **19**, 66 (2018).
9. R. Khnouf, S. Shore, C. M. Han, J. M. Henderson, S. A. Munro, A. P. McCaffrey, H. Shintaku, J. G. Santiago, Efficient production of on-target reads for small RNA sequencing of single cells using modified adapters. *Anal. Chem.* **90**, 12609–12615 (2018).
10. S. Subramanian Parimalam, Y. Oguchi, M. N. Abdelmoez, A. Tsuchida, Y. Ozaki, R. Yokokawa, H. Kotera, H. Shintaku, Electrical lysis and RNA extraction from single cells fixed by dithiobis(succinimidyl propionate). *Anal. Chem.* **90**, 12512–12518 (2018).
11. M. N. Abdelmoez, Y. Oguchi, Y. Ozaki, R. Yokokawa, H. Kotera, H. Shintaku, Distinct kinetics in electrophoretic extraction of cytoplasmic RNA from single cells. *Anal. Chem.* **92**, 1485–1492 (2020).
12. H. Shintaku, H. Nishikii, L. A. Marshall, H. Kotera, J. G. Santiago, On-chip separation and analysis of RNA and DNA from single cells. *Anal. Chem.* **86**, 1953–1957 (2014).
13. K. Kuriyama, H. Shintaku, J. G. Santiago, Isotachophoresis for fractionation and recovery of cytoplasmic RNA and nucleus from single cells. *Electrophoresis* **36**, 1658–1662 (2015).
14. A. Byrne, A. E. Beaudin, H. E. Olsen, M. Jain, C. Cole, T. Palmer, R. M. DuBois, E. C. Forsberg, M. Akeson, C. Vollmers, Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* **8**, (2017).
15. N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, K. Ardlie, D. A. Weitz, O. Rozenblatt-Rosen, F. Zhang, A. Regev, Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
16. S. Klinge, J. L. Woolford Jr., Ribosome assembly coming into focus. *Nat. Rev. Mol. Cell Biol.* **20**, 116–131 (2019).
17. N. Battich, T. Stoeger, L. Pelkmans, Control of transcript variability in single mammalian cells. *Cell* **163**, 1596–1610 (2015).
18. K. Bahar Halpern, I. Caspi, D. Lemze, M. Levy, S. Landen, E. Elinav, I. Ulitsky, S. Itzkovitz, Nuclear retention of mRNA in mammalian tissues. *Cell Rep.* **13**, 2653–2662 (2015).
19. M. M. K. Hansen, R. V. Desai, M. L. Simpson, L. S. Weinberger, Cytoplasmic amplification of transcriptional noise generates substantial cell-to-cell variability. *Cell Syst.* **7**, 384–397.e6 (2018).
20. B. B. Lake, R. Ai, G. E. Kaeser, N. S. Salathia, Y. C. Yung, R. Liu, A. Wildberg, D. Gao, H.-L. Fung, S. Chen, R. Vijayaraghavan, J. Wong, A. Chen, X. Sheng, F. Kaper, R. Shen, M. Ronaghi, J.-B. Fan, W. Wang, J. Chun, K. Zhang, Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
21. M. Hagemann-Jensen, C. Ziegenhain, P. Chen, D. Ramsköld, G.-J. Hendriks, A. J. M. Larsson, O. R. Faridani, R. Sandberg, Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).
22. I. Gupta, P. G. Collier, B. Haase, A. Mahfouz, A. Joglekar, T. Floyd, F. Koopmans, B. Barres, A. B. Smit, S. A. Sloan, W. Luo, O. Fedrigo, M. E. Ross, H. U. Tilgner, Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018).
23. R. Volden, C. Vollmers, Highly multiplexed single-cell full-length cDNA sequencing of human immune cells with 10X genomics and R2C2. bioRxiv 2020.01.10.902361 [**Preprint**]. 11 January 2020. https://doi.org/10.1101/2020.01.10.902361.
24. K. Kuriyama, H. Shintaku, J. G. Santiago, Protocol for microfluidic system to automate the preparation and fractionation of the nucleic acids in the cytoplasm versus nuclei of single cells. *Bio-protocol* **6**, e1844 (2016).
25. J. Wang, B. Fei, Y. Zhan, R. L. Geahlen, C. Lu, Kinetics of NF-κB nucleocytoplasmic transport probed by single-cell screening without imaging. *Lab Chip* **10**, 2911–2916 (2010).
26. Y. Zhan, V. A. Martin, R. L. Geahlen, C. Lu, One-step extraction of subcellular proteins from eukaryotic cells. *Lab Chip* **10**, 2046–2048 (2010).
27. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
28. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
29. J. Westoby, M. S. Herrera, A. C. Ferguson-Smith, M. Hemberg, Simulation-based benchmarking of isoform quantification in single-cell RNA-seq. *Genome Biol.* **19**, 191 (2018).
30. A. D. Tang, C. M. Soulette, M. J. van Baren, K. Hart, E. Hrabeta-Robinson, C. J. Wu, A. N. Brooks, Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat. Commun.* **11**, 1438 (2020).
31. Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, S. K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
32. H. Shintaku, J. W. Palko, G. M. Sanders, J. G. Santiago, Increasing hybridization rate and sensitivity of bead-based assays using isotachophoresis. *Angew. Chem. Int. Ed.* **53**, 13813–13816 (2014).

Acknowledgments: We thank K. Iida of the Kyoto University and S. Uemura of the University of Tokyo for critical comments on our manuscript. **Funding:** This work was supported by the ImPACT program (CSTI, Cabinet Office, the Government of Japan), JSPS Core-to-Core Program, White Rock Foundation, and the Japan Society for the Promotion of Science (grant numbers 26289035, 19H02072, and 19 K22121). M.N.A. acknowledges the Marubun Research Promotion Foundation (Japan) for the Exchange Research grant. **Author contributions:** Y.Oz. and M.N.A. performed the experiments. Y.Og. and H.S. analyzed the data. Y.Og., M.N.A., and H.S. wrote the paper. H.S. designed and supervised the study. All authors read and approved the final manuscript. **Competing interests:** Y.Og. and H.S. are inventors on a patent related to this work filed by Kyoto University and the University of Tokyo (no. 6338262, filed 27 January 2017 and published 3 August 2017). The authors declare that they have no other competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The datasets generated in this study have been deposited in the DDBJ DRA repository under the accession number DRA010535. Additional data related to this paper may be requested from the authors.

Submitted 27 July 2020
Accepted 19 February 2021
Published 7 April 2021
10.1126/sciadv.abe0317

Citation: Y. Oguchi, Y. Ozaki, M. N. Abdelmoez, H. Shintaku, NanoSINC-seq dissects the isoform diversity in subcellular compartments of single cells. *Sci. Adv.* **7**, eabe0317 (2021).