

Electronic Supplementary Information (ESI)

μCB-seq: Microfluidic cellular barcoding and sequencing for high-resolution imaging and sequencing of single cells

Authors: Tyler N. Chen, ^{*ab} Anushka Gupta, ^{*c} Mansi Zalavadia, ^a Aaron M. Streets ^{+acd}

Author information

^a University of California, Berkeley, Department of Bioengineering, Berkeley, CA, 94720, United States of America

^b University of California, Berkeley, Department of Materials Science and Engineering, Berkeley, CA, 94720, United States of America

^c UC Berkeley-UC San Francisco Graduate Program in Bioengineering, Berkeley, CA, 94720, United States of America

^d Chan Zuckerberg Biohub, San Francisco, CA, 94158, United States of America

* These authors contributed equally to this work

+ Corresponding Author: Aaron M. Streets, PhD

Assistant Professor, Department of Bioengineering, University of California, Berkeley, California, 94720, United States of America

Tel.: (510) 842- 7873, E-mail: astreet@berkeley.edu

Funding: This publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM124916. Anushka Gupta is supported by the UC Berkeley Lloyd fellowship in Bioengineering. AMS is a Chan Zuckerberg Investigator.

Conflicts of interest: There are no conflicts to declare.

Supplemental Note 1

Imaging Chamber Volume Measurement for Trapping 10 pg of Total RNA

When measuring chamber volume for Total RNA experiments in the μ CB-seq device (Fig. 3), we observed a discrepancy in height between the μ CB-seq flow molds and the reaction chambers of the PDMS μ CB-seq devices with actuated control valves. Flow molds were measured by Dektak profilometer, giving an imaging chamber height of 29 μ m. When imaging the corresponding chamber on the μ CB-seq device via Coherent anti-Stokes Raman spectroscopy (CARS), we recorded a chamber height of 53.5 μ m. Profilometry was not feasible for the closed μ CB-seq device, so we elected to conservatively use the CARS measurement at the risk of overestimating volume and loading less than 10 pg Total RNA into the μ CB-seq device. To measure chamber volume, we pressurized the isolation valves on a μ CB-seq device and acquired a z-stack of the resultant air-filled imaging chamber. Images were thresholded in ImageJ and manually outlined to record the cross-sectional area of each imaging chamber slice. The volume of the chamber was estimated by a Riemann sum to ensure that chamber volume erred on the larger side. The chamber volume measured by this method was 1.88 nL, which resulted in our conservative input concentration of 5.31 ng/ μ L Total RNA to ensure no more than 10 pg of RNA was processed in each lane of the μ CB-seq device.

Supplemental Note 2

Comparison of μ CB-seq and Fluidigm C1 performance

The Fluidigm C1 is a commercial microfluidic platform with integrated valves that uses SMART-seq for full-length transcript quantification in single cells.¹ Arguel et al. demonstrated the use of Fluidigm C1 to sequence single HEK293T cells with a 5' UMI tagging protocol.² We used this study to benchmark the current performance of μ CB-seq against the Fluidigm C1, as both studies process HEK293T cells and implement UMI-based transcript counting. There are some limitations to this comparison, however. First, there may be differences in capture efficiency and bias between a 5' sequencing chemistry and the 3' sequencing chemistry used here. Additionally, Arguel *et al.* used the hg19 genome for alignment with STAR, whereas we used the GRCh38 genome with STAR. Finally, Arguel et al. used Dropseq Core³ for UMI counting whereas we used zUMIs⁴ and filtered for exons. Using 500k reads as an individual point of comparison, the protocol on the C1 detected a mean of ~6,000 genes. Another published result using the standard SMARTer protocol on the C1 suggests similar gene detection level of ~6,000 genes, although

this was carried out on a different cell type (HCT116).⁵ Our μ CB-seq protocol detected a mean of 6,663 genes when counting only exons, and 9,203 genes when counting exons and introns (Fig. S12). This suggests that μ CB-seq has similar or improved sensitivity compared to other protocols on the Fluidigm C1.

With regards to sample-to-sample variability, Arguel et al. compute pairwise correlations across different microfluidic runs (with 37, 47 or 74 cells). For each device run, log-transformed mean UMI counts across all the cells were used as input for correlation analysis, giving $R = 0.92, 0.98,$ and 0.93 . Our μ CB-seq device has a slightly lower but comparable correlation value of $R = 0.91$ (p -value < 0.05) between device runs with 8 and 7 cells (Fig. S13). This slightly lower correlation value is expected due to the 4- to 9-fold higher number of cells averaged in the C1 experiments as compared to our μ CB-seq experiments.

Supplemental Figures

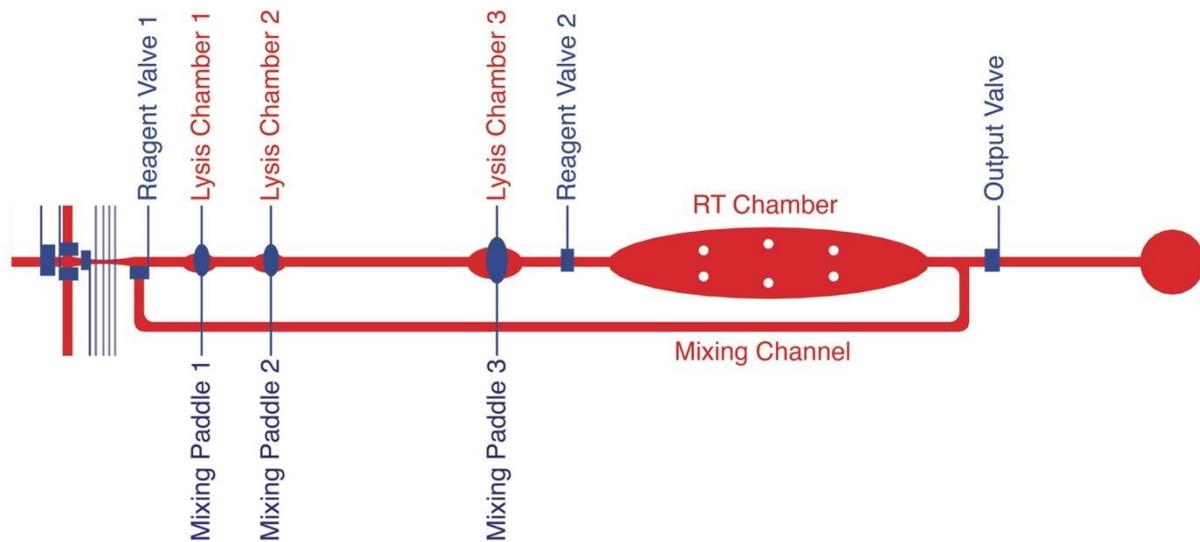


Figure S1 Detailed schematic of a single reaction lane on the μ CB-seq device. The lysis module has 3 reaction chambers and the RT module has 1 reaction chamber connected to the mixing channel. Both lysis and RT modules are separated from each other by the two reagent valves. RT primers with known barcode sequences are spotted in the *Lysis Chamber 3* of each reaction lane. Positioned atop each of the reaction chambers in the lysis module are mixing paddles, which are actuated to resuspend barcoded RT primers in lysis buffer and circulate the relatively viscous RT mix throughout the mixing channel.

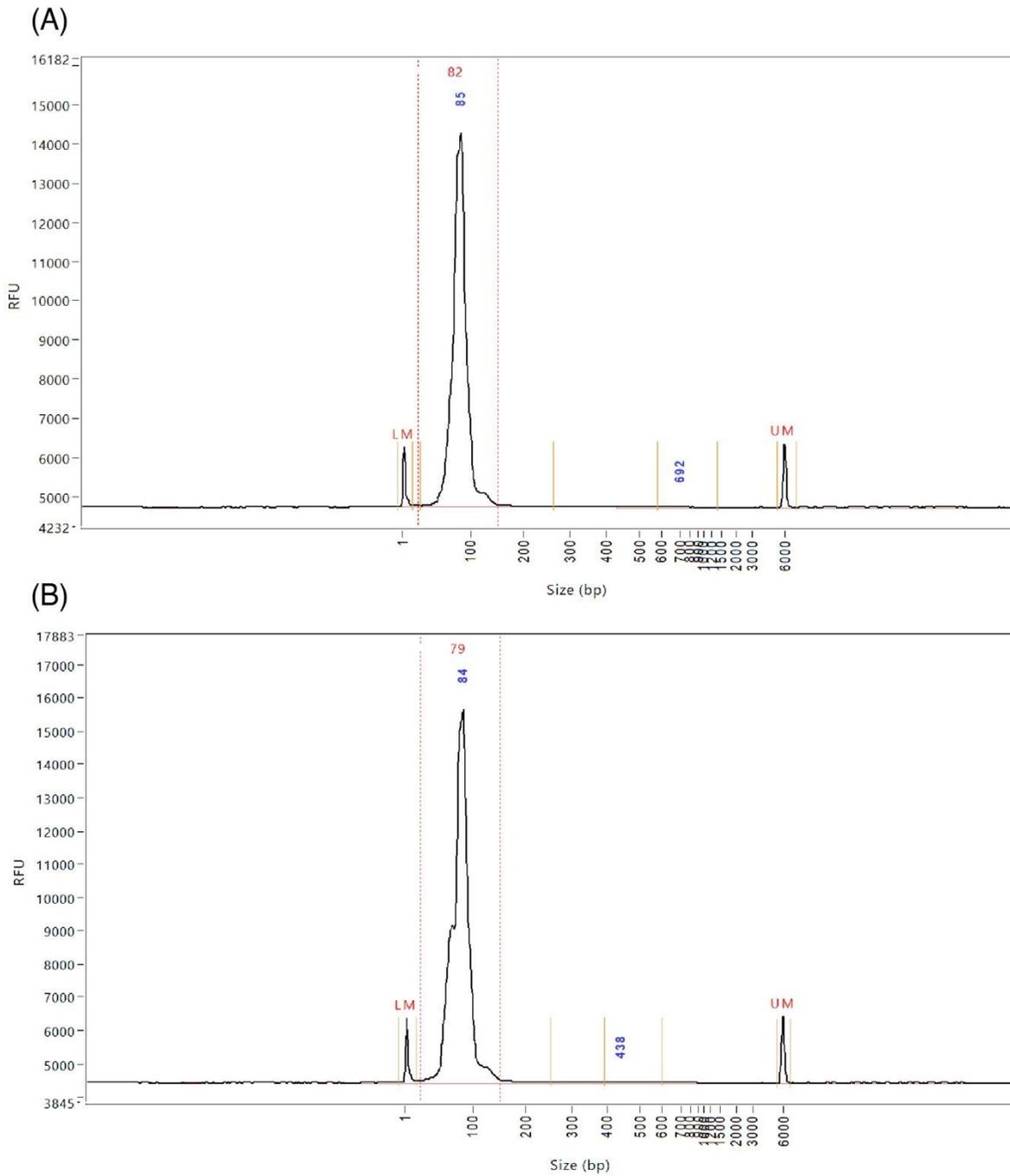


Figure S2 Validation of intact RT primer recovery from a PDMS slab after baking.

Fragment analysis size distribution traces for barcoded primers that were suspended in nuclease-free water at RT and (A) left in the original tube or (B) spotted on PDMS, dried, baked at 80 °C and recovered by resuspending in nuclease-free water.

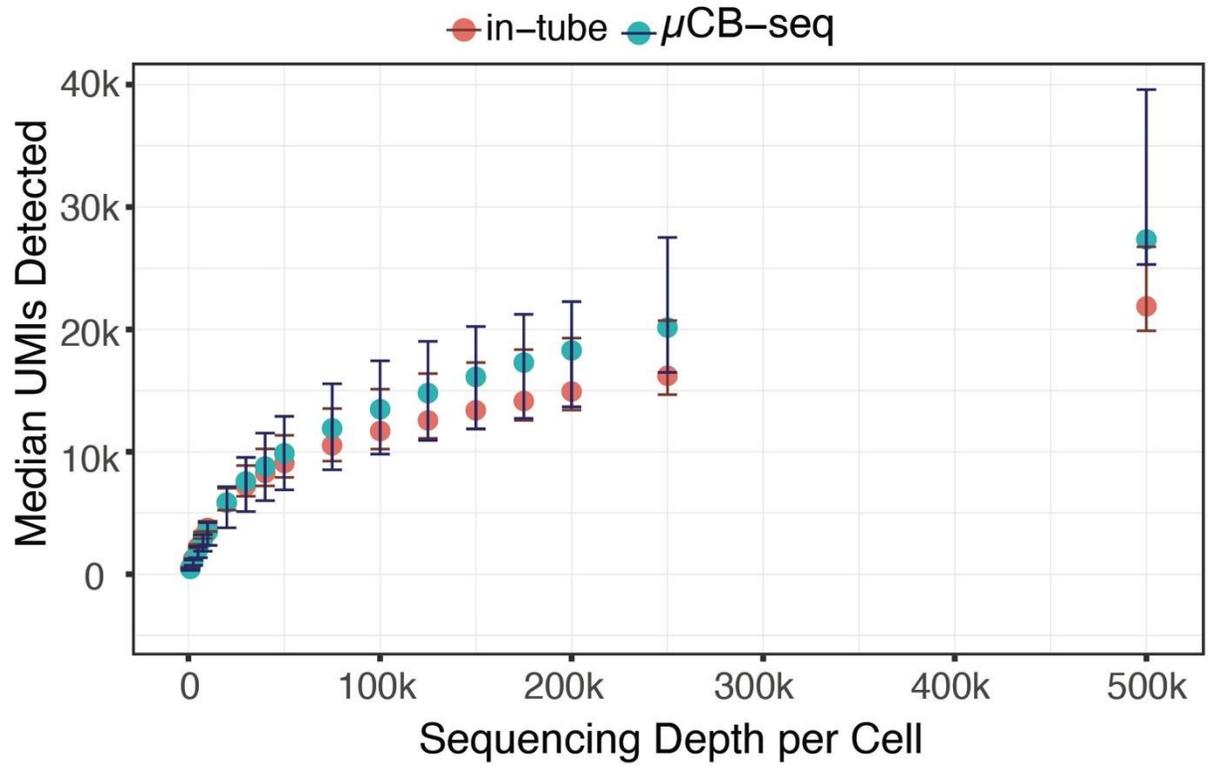


Figure S3 Median UMIs detected for downsampled read depth across single HEK cells sequenced using μ CB-seq (n = 16) and mcSCRB-seq in-tube (n = 16). Error bars indicate the interquartile range.

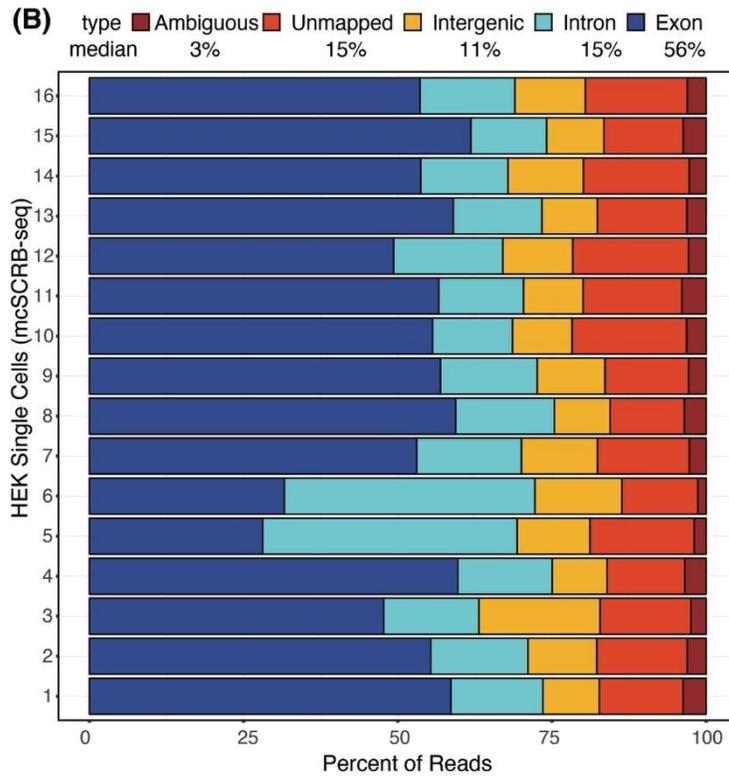
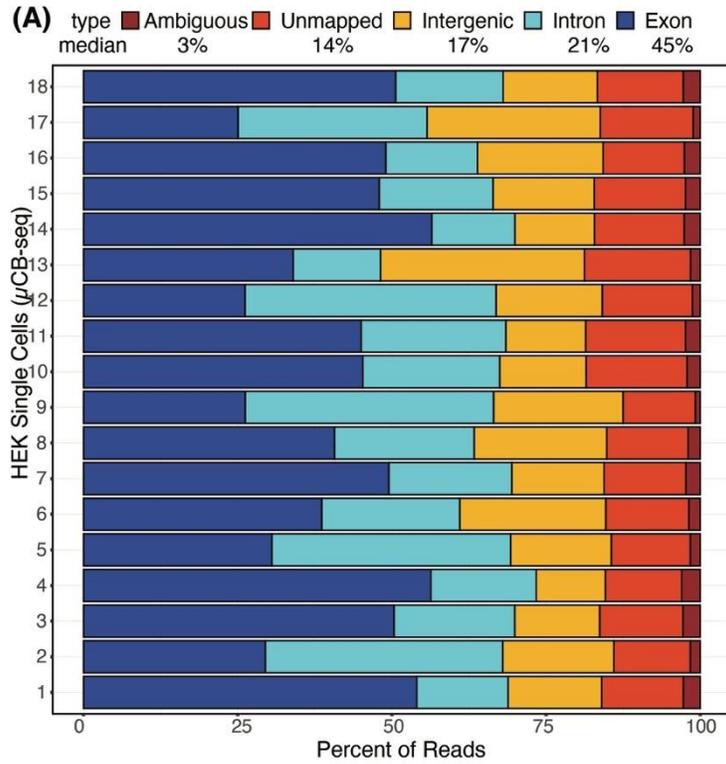


Figure S4 Mapping Statistics for single HEK Cells sequenced using (A) μ CB-seq and (B) mcSCRIB-seq in-tube.

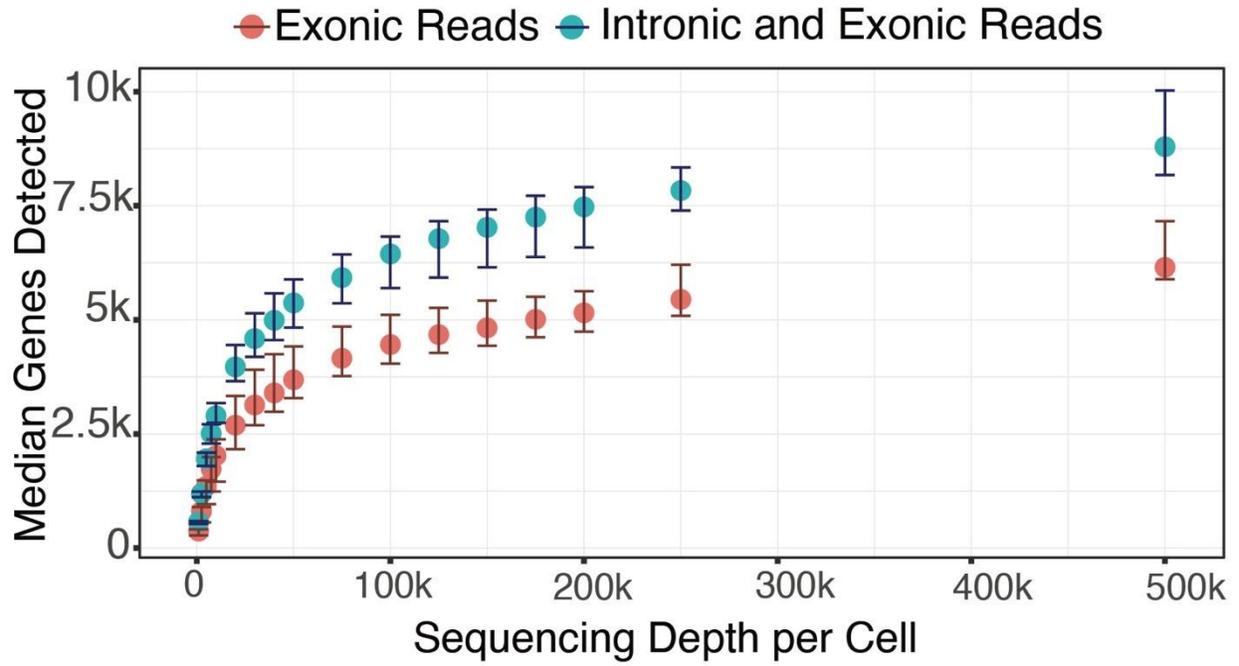


Figure S5 Median genes detected using only exonic or both exonic and intronic reads for downsampled read depths across single HEK cells (n=16) sequenced using μ CB-seq. Error bars indicate the interquartile range.

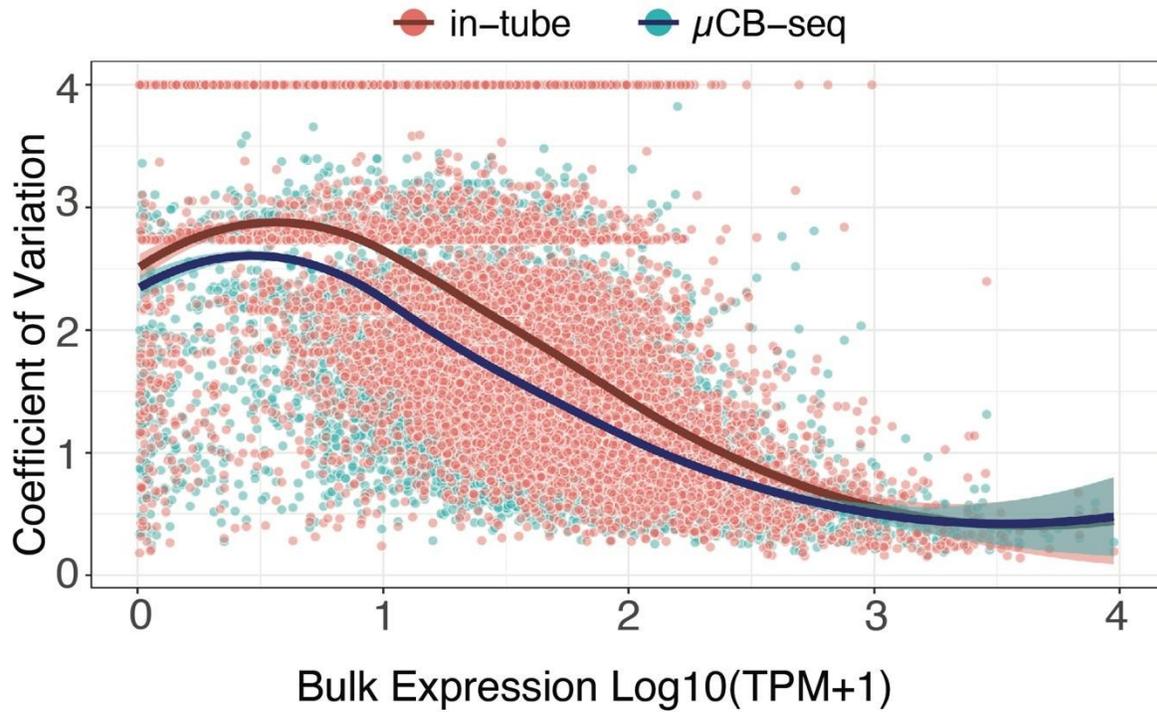


Figure S6 μ CB-seq and in-tube mcSCRB-seq protocol have comparable precision. The coefficient of variation for each gene (SD normalized by the mean) is plotted against its bulk expression for HEK cells sequenced using μ CB-seq (n=16) and mcSCRB-seq in-tube (n=16). HEK Cells were sequenced to a depth of 200,000 reads and bulk RNA-seq library was prepared using 1ug HEK total RNA sequenced to a depth of 63 million reads. CV was calculated for all common genes detected in bulk RNA-seq, μ CB-seq -seq and mcSCRB-seq libraries. The highlighted region displays the 95% confidence interval around the smooth fit as determined by loess regression.

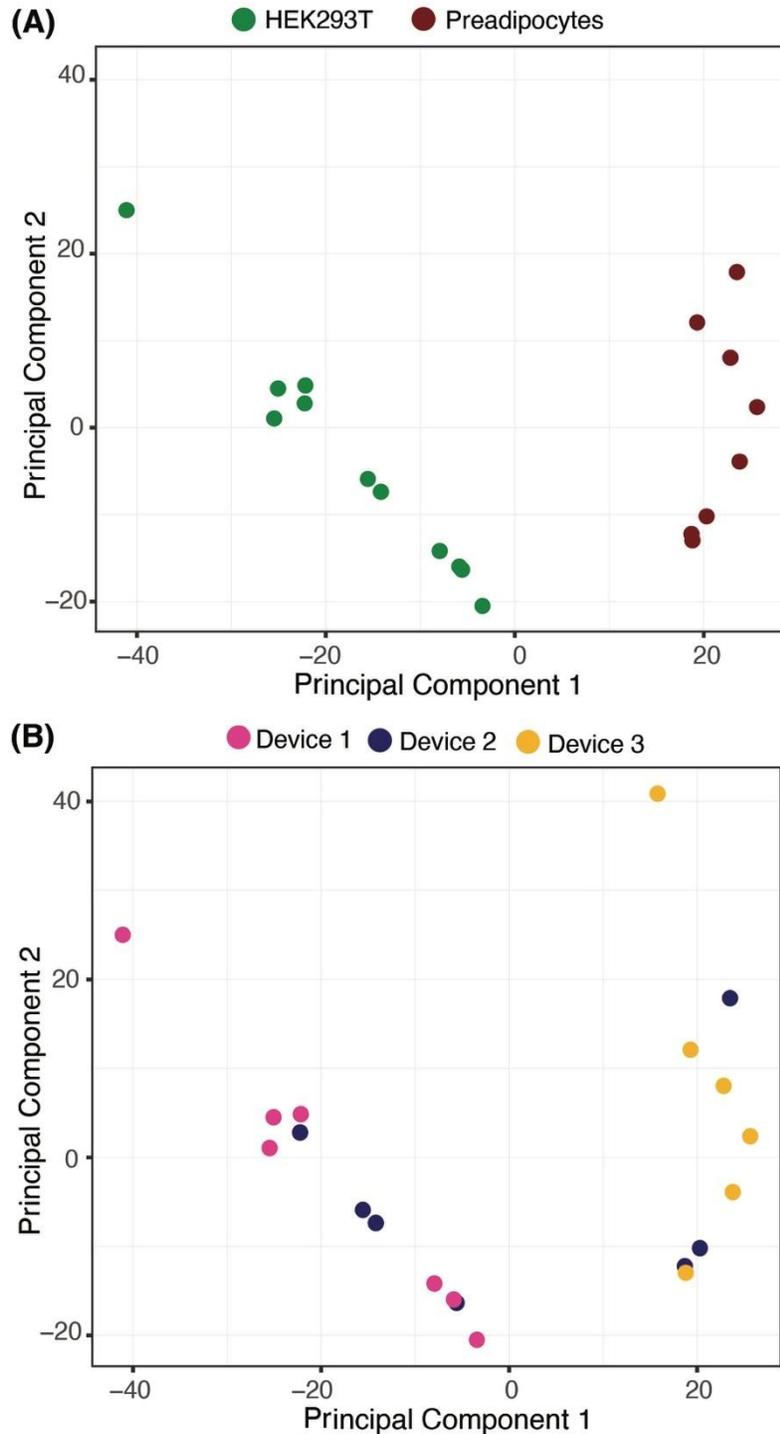


Figure S7 Annotation of HEK293T cells and Preadipocytes in Principal Component Space based on (A) Cell-clusters identified using unsupervised hierarchical clustering in the PCA space and (B) μ CB-seq devices on which cells were processed. Device 1 processed just HEKs (n=7), Device 2 processed a mix of both HEKs (n=4) and preadipocytes (n=3), and Device 3 processed just Preadipocytes (n=6)

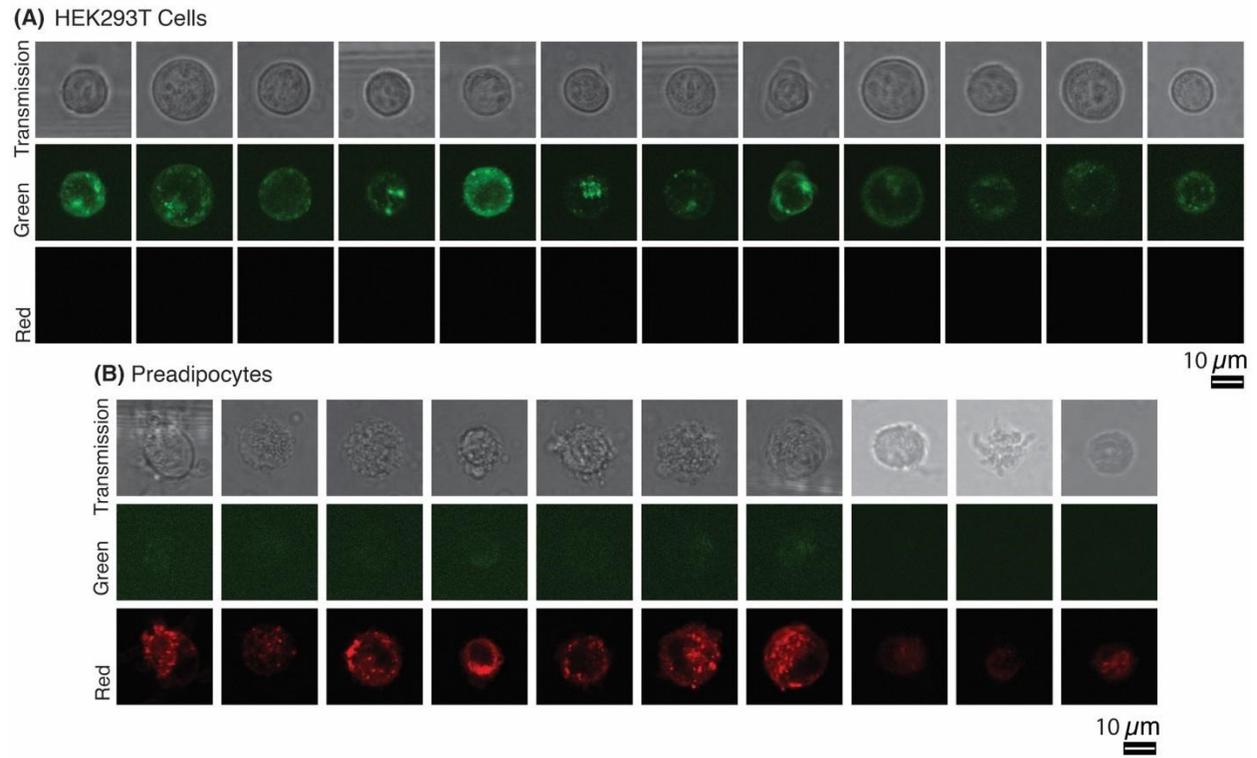


Figure S8 Scanning transmission and two-channel fluorescent confocal images of all (A) HEK293T cells and (B) Preadipocytes stained using CellBrite™ Green and Red dye respectively.

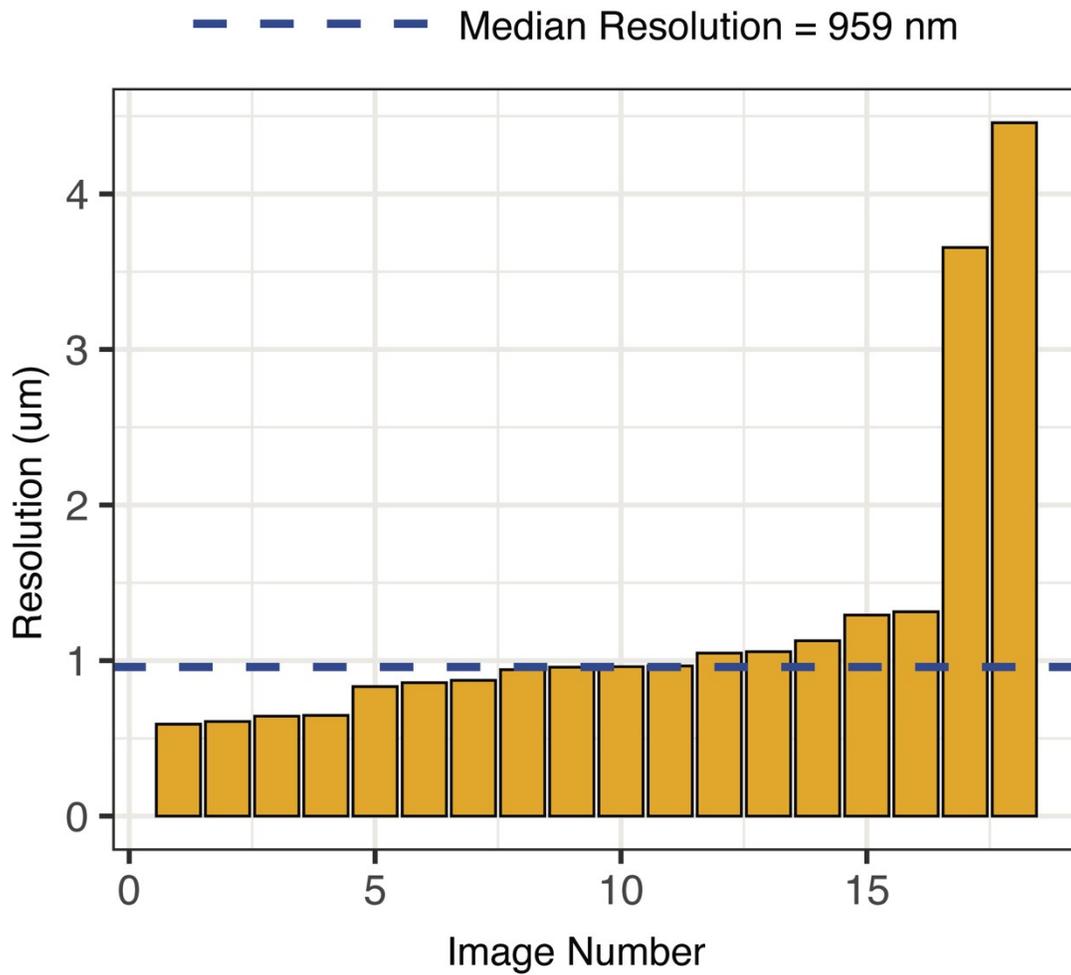


Figure S9 Spatial resolution in confocal fluorescent images of HEK cells and Preadipocytes. The blue dashed line indicates the median resolution across 18 images. Detailed image analysis steps are explained in the *Materials and Methods* section.

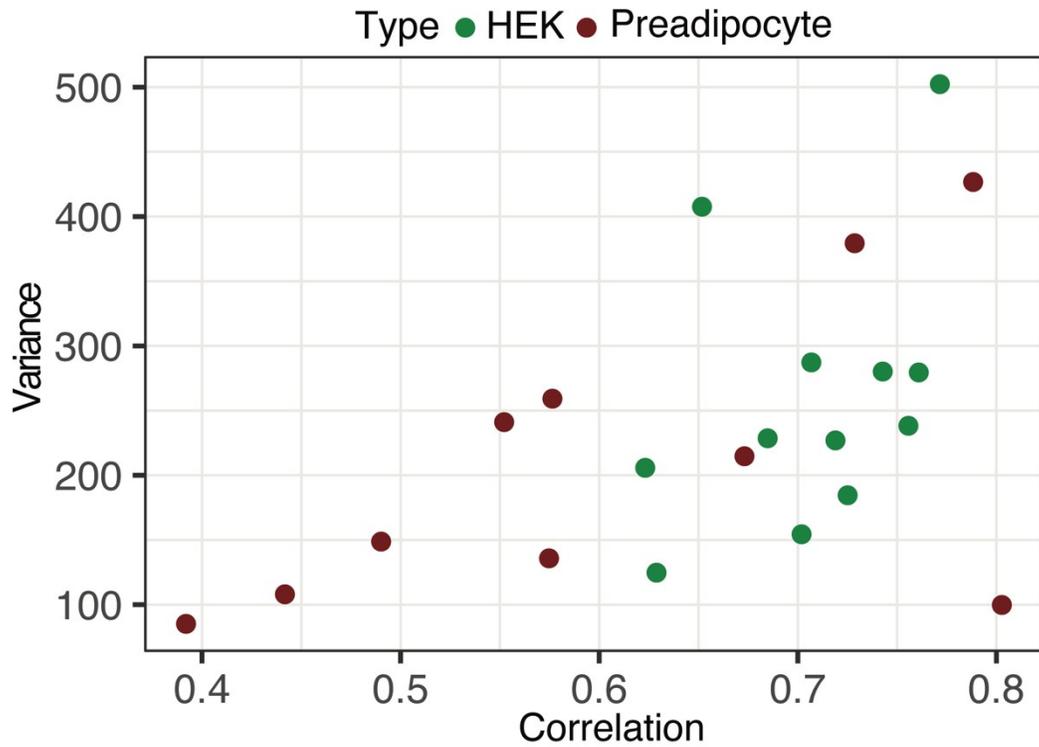


Figure S10 Annotation of HEK293T cells and Preadipocytes in a 2-Dimensional Correlation vs Variance space as quantified for grayscale intensities in the scanning transmission images.

Detailed image analysis steps are explained in the *Materials and Methods* section.

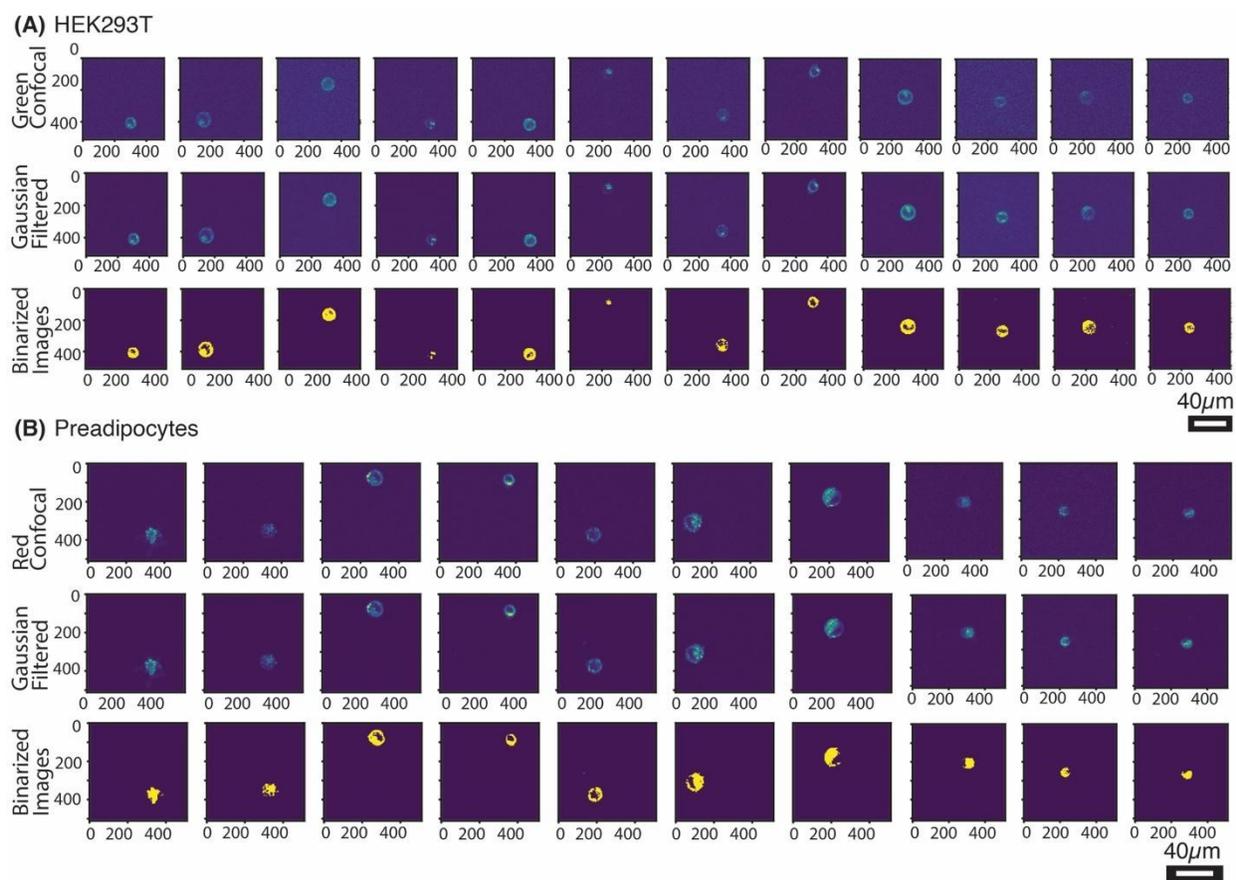


Figure S11 Fluorescent confocal, Gaussian filtered and Otsu thresholded images of all (A) HEK293T cells and (B) Preadipocytes analyzed in Figure 5 in the main text. Detailed image analysis steps are explained in the *Materials and Methods* section.

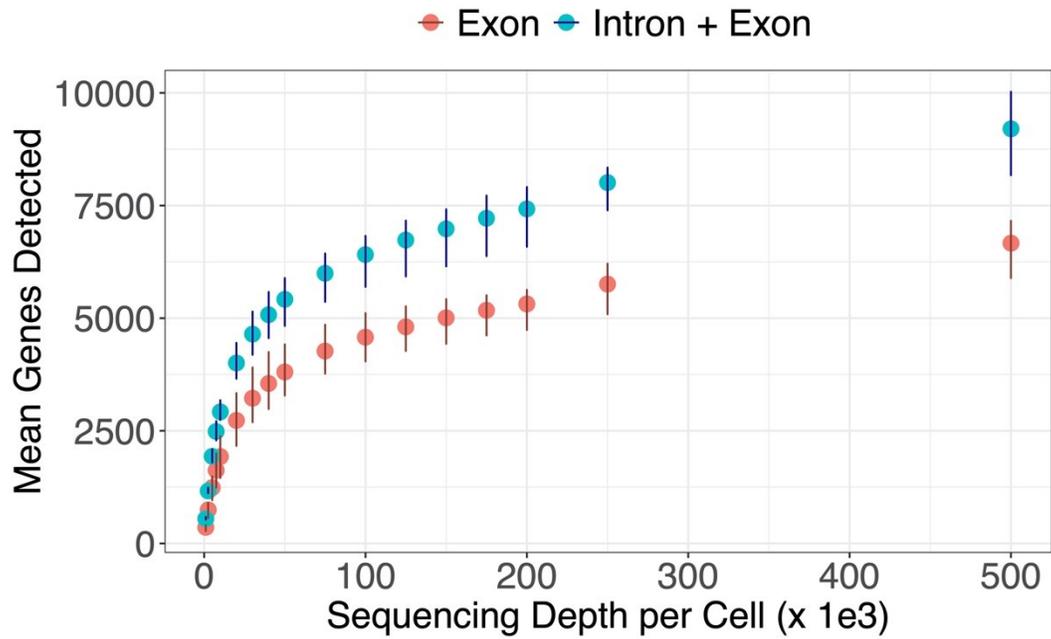


Figure S12 Mean genes detected using only exonic or both exonic and intronic reads for downsampled read depths across single HEK cells (n=16) sequenced using μ CB-seq. Error bars indicate the interquartile range.

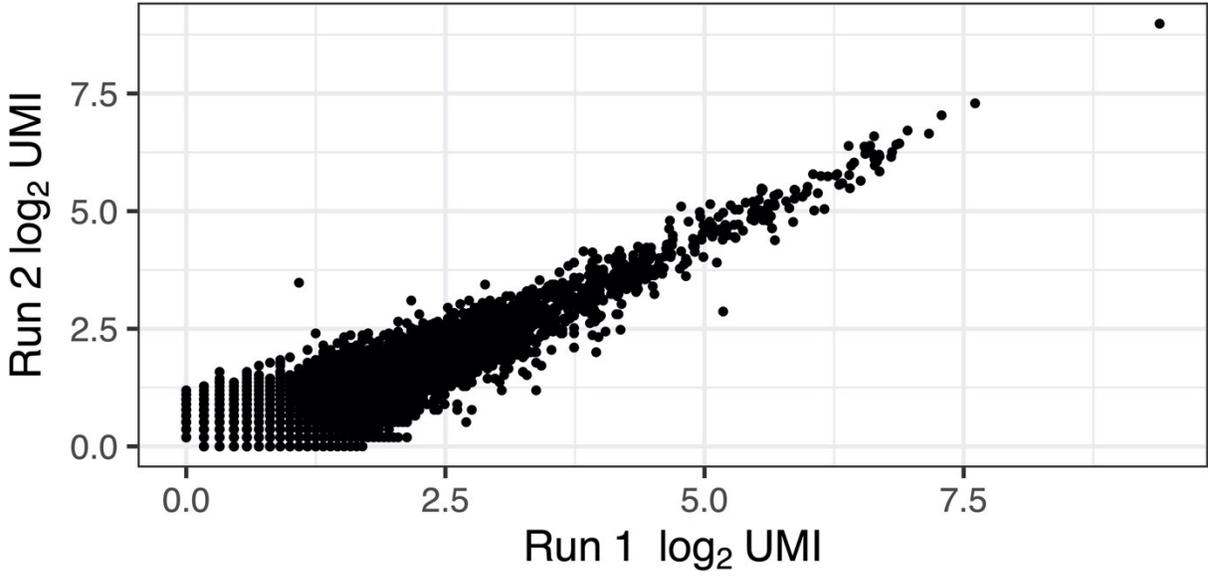


Figure S13 Pairwise correlation of the mean UMI transcript counts for two μ CB-seq HEK293T single cell transcriptome sequencing experiments. Each dot represents the log-transformed mean UMI counts for a given transcript for all cells at a depth of 250,000 reads per cell. Data from 8 and 7 cells are shown for Run1 and Run2 respectively.

<u>bc13</u>	TGCTACAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCTACAG NNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
<u>bc15</u>	CGCTATGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTATGA NNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
<u>bc26</u>	ATGCACGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCACGT NNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
<u>bc40</u>	TATGCACG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTATGCACG NNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
<u>bc47</u>	CATCGTGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATCGTGA NNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
<u>bc82</u>	CCAGTTAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAGTTAG NNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
<u>bc92</u>	GGCATTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCATTGT NNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN

Table S2 Sequences of DNA primers used in both mcSCRB-seq in-tube experiments and on μ CB-seq devices for off-chip library preparation. Same primer sequences as in mcSCRB-seq⁷ are used in this work. /5Biosg/ indicates a 5' Biotin, *_ indicates a phosphorothioated nucleotide, and r_ indicates an RNA base.

Primer	Sequence
SINGV6	/5Biosg/ACACTCTTTCCCTACACGACGC
P5NEXTPT5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCG*A*T*C*T
E5V6 TSO	CGCACACTCTTTCCCTACACGACGCrGrGrG

References

- 1 A. A. Pollen, T. J. Nowakowski, J. Shuga, X. Wang, A. A. Leyrat, J. H. Lui, N. Li, L. Szpankowski, B. Fowler, P. Chen, N. Ramalingam, G. Sun, M. Thu, M. Norris, R. Lebofsky, D. Toppani, D. W. Kemp, M. Wong, B. Clerkson, B. N. Jones, S. Wu, L. Knutsson, B. Alvarado, J. Wang, L. S. Weaver, A. P. May, R. C. Jones, M. A. Unger, A. R. Kriegstein and J. A. A. West, *Nat. Biotechnol.*, 2014, **32**, 1053–1058.
- 2 M. J. Arguel, K. Lebrigand, A. Paquet, S. R. García, L. E. Zaragosi, P. Barbry and R. Waldmann, *Nucleic Acids Res.*, 2017, **45**, e48.
- 3 E. Z. Macosko, A. Basu, R. Satija, J. Nemes, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev and S. A. McCarroll, *Cell*, 2015, **161**, 1202–1214.
- 4 S. Parekh, C. Ziegenhain, B. Vieth, W. Enard and I. Hellmann, *Gigascience*, 2018, **7**.
- 5 A. R. Wu, N. F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M. E. Rothenberg, F. M. Mburu, G. L. Mantalas, S. Sim, M. F. Clarke and S. R. Quake, *Nat. Methods*, 2014, **11**, 41-46.
- 6 L. V. Bystrykh, *PLoS One*, 2012, **7**, e36852.
- 7 J. W. Bagnoli, C. Ziegenhain, A. Janjic, L. E. Wange, B. Vieth, S. Parekh, J. Geuder, I. Hellmann and W. Enard, *Nat. Commun.*, 2018, **9**, 2937.