RESEARCH ARTICLE SUMMARY

RESEARCH METHODS

Multistep genomics on single cells and live cultures in subnanoliter capsules

Ignas Mazelis, Haoxiang Sun, Arpita Kulkarni, Theresa L. Torre, Allon M. Klein*

Full article and list of author affiliations: https://doi.org/10.1126/science.ady7209

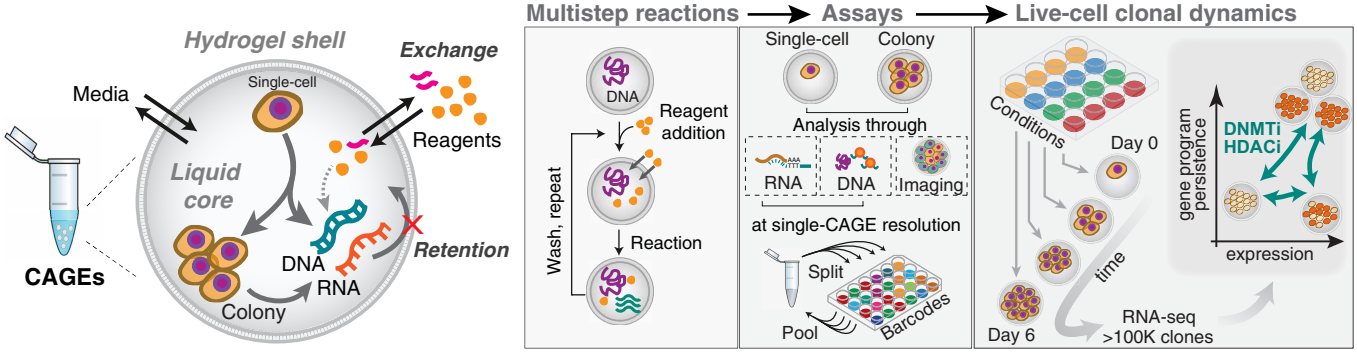
INTRODUCTION: Single-cell genomic technologies have transformed our ability to measure molecular states across thousands to millions of cells in parallel. Yet these approaches are inherently destructive, making it difficult to connect live-cell behaviors with subsequent molecular readouts. By contrast, assays on living cells can reveal dynamic behaviors and causal mechanisms, but they are limited in their molecular breadth. As a result, many dynamic properties of cells, such as the persistence of epigenetic memory or the response of cells to ligands, remain hard to analyze systematically at scale. Resolving this constraint requires a scalable way of compartmentalizing cells, which would support live-cell culture and sequential reactions.

RATIONALE: We set out to establish semipermeable capsules as a means to bridge high-throughput genomic analysis with assays on living cells. These capsules are subnanoliter aqueous compartments that resemble droplet emulsions but are enclosed by a thin hydrogel shell. The shell permits diffusion of small molecules and enzymes while retaining cells and larger macromolecules, allowing reagents to be exchanged without loss of material. To be useful, these capsules should allow live-cell growth, have appropriate permeability, and have the capacity to be barcoded at scale to connect genomic readouts to individual capsules. Together, these capabilities would provide a basis for multistep genomic workflows and dynamic measurements on living cells.

RESULTS: We identified several amphiphilic block copolymers as biocompatible materials that can form robust capsules. These capsules with amphiphilic gel envelopes (CAGEs) retain nucleic acids and other macromolecules while allowing rapid diffusion of media, enzymes, and primers. They are both permeable and stable, enabling sequential molecular reactions with repeated addition and removal of buffers, enzymes, and cofactors within intact capsules, and allowing multistep genomic workflows without material loss. We established a split-and-pool barcoding approach that assigns a unique DNA barcode identifier to each CAGE, enabling high-throughput single-cell assays. These compartments also support long-term clonal expansion of diverse cell types, including human induced pluripotent stem cells and mouse hematopoietic stem cells. Using in-CAGE barcoding followed by RNA sequencing, we quantified the persistence of transcriptional programs in tens of thousands of expanding cancer cell colonies and their response to drugs that target enzymes implicated in maintaining epigenetic gene silencing. In these assays, we increased the number of colonies analyzed by about 1000-fold over previous studies carried out in wells. We found that the drugs we tested do not uniformly activate gene expression or reduce the persistence of epigenetic states. Instead, they destabilize specific programs while stabilizing others.

CONCLUSION: Semipermeable capsules, such as CAGEs, are a versatile tool for high-throughput molecular processing of cells and molecules. They allow live-cell growth to be combined with subsequent genome-scale analysis. Their selective permeability, mechanical robustness, and scalable barcoding enable complex, multistep workflows that were previously impractical. CAGEs offer a foundation for linking molecular and functional phenotypes and for expanding genomic analysis to time-dependent processes such as differentiation, adaptation, and drug response. □

*Corresponding author. Email: allon_klein@hms.harvard.edu Cite this article as I. Mazelis *et al.*, *Science* **391**, 1130 (2026). DOI: 10.1126/science.ady7209



Compartments for multistep analysis of single cells and live cultures. CAGEs retain cells and macromolecules while freely exchanging reagents, such as enzymes, primers, and media. They allow multistep reactions, thus enabling composite genomic assays. By culturing cells in CAGEs before genomic profiling, this study measures the persistence of gene programs and changes in persistence in response to certain epigenetic drugs. DNMTi, DNA methyltransferase inhibitor; HDACi, histone deacetylase inhibitor.

RESEARCH METHODS

Multistep genomics on single cells and live cultures in subnanoliter capsules

Ignas Mazelis[1], Haoxiang Sun[1], Arpita Kulkarni[2], Theresa L. Torre[1], Allon M. Klein[1]*

Single-cell sequencing methods uncover natural and induced variation between cells. Many functional genomic methods, however, require multiple steps that cannot yet be scaled to high throughput, including assays on living cells. In this study, we developed capsules with amphiphilic gel envelopes (CAGEs), which selectively retain cells and large analytes while being freely accessible to media, enzymes, and reagents. Capsules enable high-throughput multistep assays combining live-cell culture with genome-wide readouts. We established methods for barcoding CAGE DNA libraries and applied them to measure persistence of gene expression programs in cells by capturing the transcriptomes of tens of thousands of expanding clones in CAGEs. The compatibility of CAGEs with diverse enzymatic reactions will facilitate the expansion of the current repertoire of single-cell high-throughput measurements and their extension to live-cell assays.

Single-cell genomics encompasses a set of methods whereby hundreds to millions of cells are individually subjected to highly multiplexed assays, including DNA sequencing, chromatin accessibility or modification, RNA sequencing, and combinations thereof (*1*, *2*). These methods enable unbiased, systematic discovery of cellular phenotypes and their dynamics, including immune cell responses, transitional states in tissue development, and changes in disease or upon perturbation. However, integrating these tools with each other or with functional cell screens faces limitations because the amount of material that can be obtained from each single cell is small and often requires reaction conditions incompatible with each other (*1*). Balancing sensitivity, scale, and genome-wide breadth with live-culture capabilities imposes trade-offs between assay type, performance, and scalability.

Single-cell sequencing methods are currently implemented through three technological modalities: in wells [e.g., switch mechanism at the 5′ end of RNA templates (SMART-seq3)] (*3*, *4*), in microcompartments (droplets or nanowells) (*5*–*7*), or by in-cell barcoding (ICB) of cross-linked cells [single-cell combinatorial indexing RNA sequencing (sciRNA-seq) and split-pool ligation-based transcriptome sequencing (SPLiT-seq)] (*8*, *9*). These methods each offer different trade-offs between sensitivity and number of cells that can be analyzed, and each suffer different limitations regarding the complexity of assays they enable. Well-based methods are typically the most sensitive, possibly because each cell is processed in multiple steps in a manner resembling bulk methods. These methods also enable carrying out functional or live-imaging assays on living cells before genomic analysis. However, well-based methods scale poorly. Microcompartment-based methods achieve increased throughput by physically isolating individual cells together with individual DNA-barcoded beads, enabling cell-specific barcoding of material from thousands of cells together within one sample. However, these methods are limited to a single reaction step once cells are compartmentalized. ICB methods scale to exponentially large numbers of cells but require cell cross-linking, which leads to material loss and reduced measurement sensitivity (*8*, *9*). Neither droplets nor ICB methods allow parallelizing functional live-cell assays. Accordingly, approaches have been developed to preprocess cells before their analysis by these modalities, such as the single-cell enzyme-linked immunosorbent assay (ELISA) (*10*, *11*) or DNA tagging of cell populations (*12*, *13*).

A technology that may unify advantages of these modalities is semipermeable capsules (*14*, *15*). Capsules are subnanoliter spherical liquid compartments surrounded by a resilient shell. The chemical composition of the shell can be chosen to be selectively permeable to small molecules, including tissue culture media, small proteins such as DNA polymerases and ligases, and oligonucleotides, while selectively retaining larger macromolecules such as DNA or mRNA (Fig. 1A). Like tissue-culture wells, they have the potential to allow culturing cells in isolation before analysis, with free exchange of media and nutrients. They also allow carrying out optimized molecular biology reactions on compartmentalized cells or analytes by enabling the repeated exchange of buffers and reagents. Like droplets, capsules can be produced at scale. Their resilient shell allows sorting to enrich cells or colonies with desired features (*15*). These properties expand the range of genomic assays that are feasible for single cells and live-cell cultures and present a versatile technological platform that may allow for the parallelization of assays that, until now, have been difficult and costly to carry out at high throughput (Fig. 1B). Here, we report the development of capsules with appropriate permeability and stability for genomic assays. We implemented a split-pool barcoding-based approach to generate single-capsule genomic and transcriptomic libraries, and we applied these developments to perform a massively parallel live-cell assay characterizing the diversity and persistence of transcriptome-wide expression fluctuations in clones. The utility of capsules for multistep genomics is also shown in an accompanying study (*16*) exploring different applications.

Developing capsules with amphiphilic gel envelopes (CAGEs)

Capsules are generated by forming concentric layers of core and shell components (*17*), with one approach using liquid-liquid phase-separation (LLPS) between two polymers inside microfluidic droplets (*14*, *15*). The outer polymer can then be cross-linked to form a stable shell. For biological applications, several studies have reported production of capsules consisting of a polyethylene glycol diacrylate (PEGDA) shell (*14*, *15*). However, PEGDA fails to form stable shells under changes in buffer pH, salinity, and concentration, thereby limiting the ability to control capsule permeability (*18*, *19*) (fig. S1A). Amphiphilic molecules should more reliably wet the water-oil interface of a droplet than PEG, so we speculated that such molecules would be more suitable for producing stable concentric shells through spontaneous LLPS. We tested the addition of Pluronic F127, a biocompatible amphiphilic block copolymer, and found that it substantially improved the production of PEGDA capsules (fig. S1, B and C). This led us to use Pluronic F127 functionalized with diacrylate (F127DA) directly as a shell building block, without any PEGDA. The production of Pluronic diacrylate capsules occurs through LLPS between F127DA and a core polymer, dextran, inside microfluidic droplets. Once formed, the F127DA is covalently cross-linked into a physical gel, using a photoinitiator and brief exposure to 405-nm light (Fig. 1C). The resulting capsules are optically clear (Fig. 1D), tunable by varying polymer concentrations (fig. S1D), can be reliably formed at scale (90 ± 9% recovery; table S1), are monodisperse [coefficient of variation (CV) <3%, fig. S1E], stable over time (fig. S1F), and can be used to capture cells inside during production (79 ± 6% capture rate; table S1 and movie S1). Additionally, the strategy is generalizable: Capsule production is possible with other amphiphilic copolymers of different lengths, including other Pluronics (P123DA, F68DA, and F108DA) and acrylate-polycaprolactone-PEG-polycaprolactone-acrylate (PPPDA) (Fig. 1E), enabling the formation

[1]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [2]Single Cell Core, Harvard Medical School, Boston, MA, USA. *Corresponding author. Email: allon_klein@hms.harvard.edu

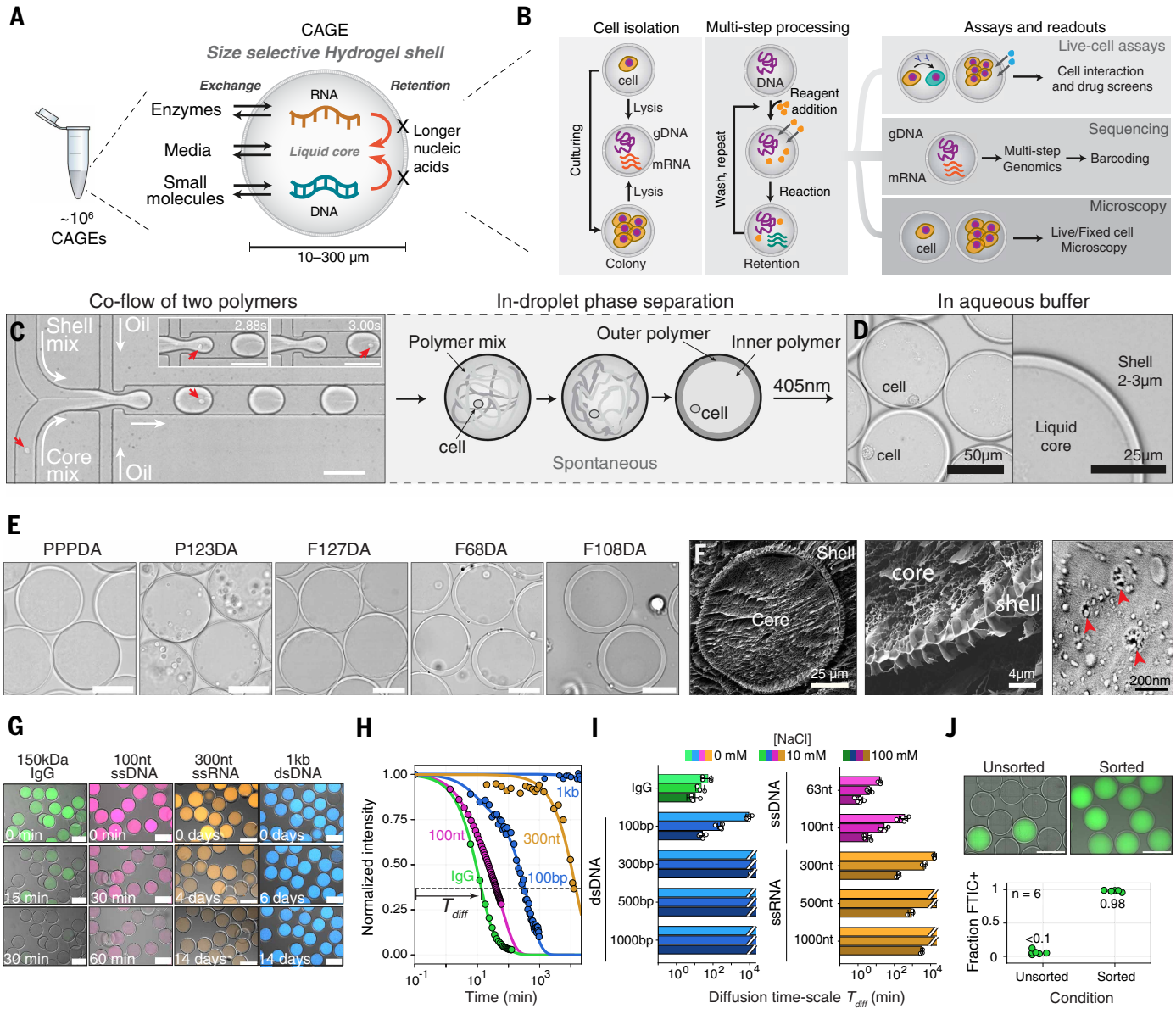


Fig. 1. Concept, development, and characterization of Pluronic diacrylate capsules. (**A**) Schematic of semipermeable capsules and their relevant properties for high-throughput, single-cell profiling. (**B**) Schematic of potential massively parallel, multistep analyses in capsules, including processing of single or cultured cells through iterative addition and removal of reagents, imaging, sorting, and sequencing. (**C**) Brightfield micrograph (left) and schematic (right) showing generation of CAGEs (compartments with amphiphilic gel envelopes) using a droplet generator coflowing an amphiphilic functionalized outer (shell) polymer and a hydrophilic core polymer that undergo phase separation upon droplet formation. These are subsequently converted to CAGEs by 405-nm–dependent cross-linking of the shell polymer. Red arrows indicate single cells. (Inset) Two snapshots spaced by 0.12 s. Scale bar, 100 μm. (**D**) Brightfield micrographs of 8% (w/v):1% (w/v) F127DA:PPPDA shell, 13% (w/v) dextran core CAGEs in Dulbecco's phosphate-buffered saline (DPBS) containing individual cells. (**E**) Brightfield micrographs demonstrating that multiple amphiphilic PEG copolymers form capsules. Scale bar, 50 μm. (**F**) Cryo-SEM images of freeze-fractured capsules with magnification of pores in the Pluronic diacrylate shell membrane. Red arrows point to surface pores. (**G**) Illustrative composite micrographs of confocal and brightfield imaging of capsules from analyte diffusion time series. IgG, immunoglobulin G. (**H**) Intensity time series quantifying diffusion half-times through the CAGE shell [composition as shown in (D)] for analytes of different sizes and in the presence of different salt concentrations. (**I**) Diffusion half-times showing a sharp size-dependent cutoff in CAGE permeability. Plot represents mean data from ≥3 independent experiments with SD as error bars; full data is found in table S2. (**J**) Flow cytometry–based enrichment of CAGEs. Composite fluorescence and brightfield micrographs of a sample initially containing <10% labeled capsules (top left) and sample after sorting (top right), yielding ~98% fluorescein isothiocyanate (FITC)–positive CAGEs (bottom). Scale bar, 100 μm.

of capsules with varying shell properties. We refer to these capsules as CAGEs: capsules with amphiphilic gel envelopes.

We next optimized the selective permeability of CAGEs. We focused on using F127DA because Pluronic F127 forms a hydrogel consisting of a packed micellar array, which would set a shell pore size comparable to the micelle diameter (radius 10 to 20 nm) (*20*, *21*). Indeed, cryo–scanning electron microscopy (cryo-SEM) imaging shows that the F127DA shell has a foam-like structure with surface pores 10 to 50 nm in diameter (Fig. 1F). Genomic assays utilize small enzymes (gyration radii 3 to 5 nm) and oligonucleotide primers of up to ~100 nucleotides (nt) in length (gyration radii <10 nm) (*22*) that should diffuse through these pores, whereas the double-stranded DNA (dsDNA) generated

during library preparation should not [persistence length >50 nm for products >150 base pairs (bp) (*23*)]. We generated CAGEs with a shell composed of F127DA and other copolymers and measured mean lifetime (T_{Diff}) associated with the diffusion of fluorescently labeled proteins, DNA, and RNA through the CAGE shell (Fig. 1G and fig. S1, G to I). We identified a polymer composition for CAGE formation for which dsDNA of length ≥300 bp showed no diffusion through the CAGE shell (diffusion lifetime T_{Diff} >14 days) in any buffer tested, whereas immunoglobulins and 100-nt single-stranded DNA (ssDNA) were able to diffuse into the CAGEs (T_{Diff} ~ 5 min) (Fig. 1, H and I, and table S2). The diffusion time remained stable over at least 10 days and after centrifugation and incubation at different temperatures (fig. S1, J and K). The diffusion rate across the shell slowed with decreasing salt concentration (Fig. 1I) and with increasing temperature (fig. S1L). This makes F127DA CAGEs particularly suitable for high-temperature reactions, such as polymerase chain reaction (PCR) (fig. S1G), in which retention of molecules at high temperature is desirable. The dependence of diffusion on salt concentration also suggests a strategy for modifying buffers for washing, storing, or loading CAGEs. Furthermore, F127DA offers a stable base for capsule formation, allowing diffusion rates to be tuned by altering polymer composition (fig. S1M).

CAGEs can be handled with common lab equipment, analyzed by flow cytometry, and sorted through a commercial fluorescence-activated cell sorting (FACS) instrument to selectively fractionate them according to fluorescence (Fig. 1J). Taken together, these results demonstrated that CAGEs offer a physically robust platform with a tunable permeability threshold. They enable exchange of oligonucleotides and proteins while retaining large macromolecules, making them suitable for multistep molecular biology reactions and thus high-throughput genomics.

Barcoding DNA in CAGEs

We reasoned that CAGEs could enable massively parallel analyses by first compartmentalizing live cells or macromolecules and subsequently performing sequential enzymatic reactions. This approach would require labeling the analytes in each compartment through DNA barcoding that later identifies their capsule of origin by sequencing.

Using CAGEs, we developed a flexible split-and-pool barcoding (*8*, *9*) approach that allows indexing of dsDNA molecules—the final form of all major analytical libraries analyzed by sequencing (Fig. 2A). The approach first amplifies the DNA derived from different assay reactions, such as reverse transcription (RT) or tagmentation, to overcome inefficiencies in subsequent barcoding. This amplification is carried out by PCR with a primer that contains a deoxyuridine (dU) residue, which can be efficiently cleaved to generate a 3′ overhang and a 5′ phosphate for subsequent barcoding (Fig. 2B and fig. S2A). This process can be iteratively repeated after each barcode addition to remove linker DNA sequences and to generate a 5′-P ligation site for the next round. In this approach, CAGEs are subject to two rounds of split-pool ligation and one round of PCR DNA barcoding (schematic in Fig. 2A), after which the prepared libraries are readily released by dissolving the CAGEs (fig. S2, B and C). Additional barcodes can be added during initial library construction (through PCR, RT, or tagmentation), as well

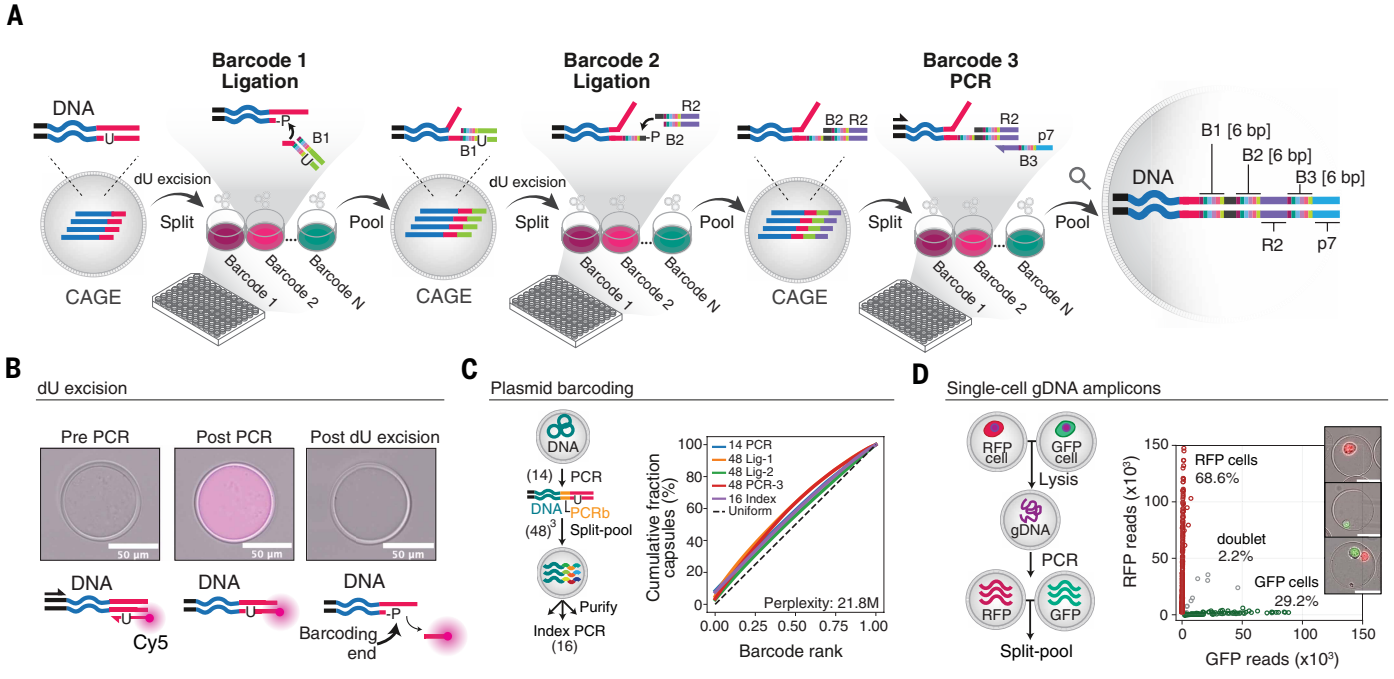


Fig. 2. Split-and-pool–based barcoding of CAGEs for high-throughput sequencing (inC-seq). (**A**) Schematic for split-and-pool barcoding of DNA libraries in capsules, showing an example of three rounds of barcoding through ligation and PCR used in this work. Initial libraries are generated with a uracil (U) residue in the common library end sequence (red), which is cleaved by a uracil-specific excision reagent to generate a sticky end with a 5′-phosphate (P). The sticky ends are ligated to an array of barcoding dsDNA oligos that contain a well-specific barcode (B1). The process is repeated to introduce a second barcode (B2), and a third barcode (B3) is introduced through a PCR primer (R2-B3-p7). (**B**) Demonstration of efficient sequential enzymatic processing of DNA in CAGEs, realizing step (1) from (A). An initial unlabeled library (left panel) is amplified by incubation of capsules in a PCR reaction mix, with a PCR primer containing a uracil and a cyanine 5 (Cy5) label at its 3′ terminus. After PCR and washing (middle panel), the CAGEs show retention of Cy5 fluorescence that now is attached to the DNA in the capsule. Uracil excision and washing cleaves the labeled primer generating an overhang and losing Cy5 fluorescence (right panel). (**C**) Representation of barcodes in the DNA amplicon sequencing data after a five-step barcoding approach: 14 PCR barcodes (PCRbs) were added during amplification, followed by 3-steps of 48 split-pool barcode addition and 16 indices following capsule solubilization. Lig, ligation barcoding. (**D**) Effective library partitioning by capsules with low cross-contamination seen by inC-seq analysis of single-cell gDNA transgenes loci encoding GFP or RFP sequences. The plot shows the number of reads from mapping to GFP or RFP, with each point corresponding to a unique capsule barcode. Observed doublet rate is in line with expectations (2.1%) for CAGEs generated with a mean cell loading of 0.1 cells per CAGE and an observed cell loading ratio of 70:30 RFP:GFP. (Inset) Examples of CAGEs with cells before lysis. Scale bar, 50 µm.

as indexing after release from CAGEs, scaling the number of barcodes. The number of capsules that can be profiled scales exponentially with the number of reaction steps (fig. S2, D and E). We refer to this method for indexing DNA libraries in capsules as "inC-seq."

We tested five-step inC-seq barcoding on CAGEs. We encapsulated plasmid DNA and then barcoded the CAGEs with $14 \times 48 \times 48 \times 48 \times 16 = 24.7 \times 10^6$ potential barcodes. Applying this protocol to 350 μl of packed CAGEs, we obtained 355,418 capsule barcodes (1 μl corresponding to ~1000 capsules), with each of the possible barcode elements uniformly represented across all five rounds of barcode addition, yielding an effective theoretical barcoding space (perplexity) of 21.8 million barcodes (Fig. 2C). We next tested inC-seq on single-cell–derived material by encapsulating cells carrying transgenes encoding red or green fluorescent proteins (RFPs or GFPs) into CAGEs and then amplifying these loci from genomic DNA (gDNA) (Fig. 2D and fig. S2F). The DNA libraries produced from cells demonstrated that the cells and genomic material within each CAGE were effectively compartmentalized through the entire multistep barcoding protocol, with 99 ± 5% (median ± SD) of reads from each CAGE mapping specifically to single GFP or RFP sequences, and with a doublet rate consistent with expectations for Poisson loading (Fig. 2D).

Single-cell genomic assays in CAGEs

DNA barcoding in capsules enables diverse analytical approaches to study DNA and RNA. We implemented single-capsule transcriptomics (inC-RNA-seq) and, separately, a single-capsule assay for transposase-accessible chromatin (*24*) (inC-ATAC-seq) (Fig. 3, A and B, and fig. S2E). We used a mixture of human and mouse cells to confirm the effective isolation of cellular material in each CAGE, with a cell doublet rate again limited by Poisson loading (Fig. 3, C and D, and fig. S3, A and B). For inC-ATAC-seq, we observed a strong enrichment of fragments at transcription start sites (17-fold enrichment) with low mitochondrial reads (<3%) and high fraction of reads in peaks (fig. S3, C to E). inC-ATAC-seq is not yet optimized, and sensitivity was low (median 2300 fragments per cell; fig. S3, F and G) but nonetheless identified differentially accessible regions in a mixture of three cell lines (fig. S3H).

We then focused on optimizing methods for high-sensitivity single-cell inC-RNA-seq using an optimized well-based approach (SMART-seq3) as a starting point (*3*). The flexibility of CAGEs allowed us to optimize sequential steps of lysis, RT, and preamplification in cell lines [human embryonic kidney (HEK) 293T (HEK293T) and NIH3T3], resulting in a protocol that captures as many transcripts and genes as a commercial droplet-microfluidic system (10x Genomics, GEMCode v3.1) (Fig. 3E and fig. S4, A and B). We observed high reproducibility between independently prepared and barcoded CAGEs (fig. S4C, log correlation $R > 0.99$) and between CAGEs and GEMCode (fig. S4D, log correlation $R = 0.95$). The porosity of capsules raises a question of whether smaller mRNA transcripts might be underrepresented, but notably we found no apparent bias in the length of detected mRNA transcripts in CAGEs, in line with our diffusion assessment data (fig. S4E). Furthermore, we observed comparable transcript counts and low length bias when performing inC-RNA-seq with different sample preparations: (i) cells fixed with formaldehyde for which cross-linking was reverted in CAGEs before RT (Fig. 3F and fig. S4F), (ii) material cryopreserved after cell lysis in CAGEs (Fig. 3G and fig. S4G), and (iii) cells encapsulated while still in tissue culture media (Fig. 3H and fig. S4H). Notably, culture media can inhibit RT, yet performance remained comparable to buffer-washed samples. These results indicate the versatility of CAGEs for complex assay design and suggest their utility for profiling difficult-to-process samples, such as microbes (*14*, *15*), cells from marine organisms (*25*), or clinical formaldehyde-fixed samples.

To complete benchmarking inC-scRNA-seq, we profiled human peripheral blood mononuclear cells (PBMCs), which have been frequently analyzed by other methods. Starting with CAGEs containing an estimated 60,000 PBMCs, we obtained transcriptomes for 43,665 PBMCs spanning the full landscape of PBMC states as seen by a uniform manifold approximation and projection (UMAP) plot (Fig. 3, I to K), including rare circulating $CD34^+$ progenitors, plasmoblasts, and a clear partitioning of subsets of T cells and natural killer (NK) cells. Our method captured comparable numbers of transcripts and genes as the GEMCode v3.1 and SMART-seq3 (*3*) methods (Fig. 3, L and M, and fig. S4, I and J) and considerably more transcripts than reported with split-and-pool ICB methods (*9*, *26*). Overall, inC-seq approaches provide a versatile platform for high-throughput genome-wide measurements, with the performance matching established transcriptomic profiling methods, and with the added flexibility of conducting multistep reactions.

Massively parallel live-cell assays in CAGEs

Capsules have the potential to enable highly parallel functional assays in which cells are allowed to grow and/or interact before molecular profiling. To enable cell culturing in CAGEs, we introduced dithiothreitol and bovine serum albumin to buffer against reactive oxygen species (ROS) formed during shell photopolymerization (*27*), which otherwise led to large-scale death of encapsulated cells (fig. S5, A and B). We also removed excess fluorinated oil, which acts as an oxygen reservoir. With these changes, cells captured in CAGEs are viable and clonally expand as demonstrated with several cell lines (L1210, K562, Jurkat, HEK293T, and L929) (Fig. 4, A and B), with division times matching those of cells grown in plates (Fig. 4C). Culturing in CAGEs did not induce major transcriptomic changes (log correlation $R^2 > 0.95$; <10 genes with >twofold change at 5% false discovery rate; fig. S5C) or stress (fig. S5D). CAGEs also support viable growth of primary mouse bone marrow–derived hematopoietic stem cells and human induced pluripotent stem cells, with the latter forming cyst-like structures (Fig. 4A) and maintaining expression of pluripotency markers (fig. S5, C and E). CAGEs can be produced at >2 million/hour and thus enable high-throughput expansion of isolated colonies at scale (Fig. 4B and fig. S5F), all of which can then be analyzed with inC-RNA-seq.

We utilized this ability to grow cells in CAGEs to measure persistent clonal heterogeneity in gene expression. In several cancer types, cells occupy persistent epigenetic states, which are thought to underlie nongenetic heterogeneity in drug resistance (*28*, *29*). Drugs targeting the inheritance of epigenetic modifications [DNA methyltransferase (DNMT) inhibitors (*30*) and histone deacetylase (HDAC) inhibitors (*31*)] have been proposed as tools to activate tumor suppressor genes that are silenced, and they are used in combination with more traditional chemotherapies (*32*–*34*). Whether these drugs broadly reduce epigenetic persistence in gene expression, however, has not been explicitly tested. To measure the duration of epigenetic persistence, other studies have used RNA-seq to identify clonally variable gene expression patterns—an approach so far carried out by culturing isolated cells in wells, with analysis restricted to a few dozen clones (*35*, *36*). Using CAGEs, one can parallelize this live-cell assay to carry out tens of thousands of clonal growth experiments. We did so with cell clones derived from a mixture of human erythroleukemia (K562) and mouse lymphoblastoma (L1210) cells and evaluated changes in clonal heterogeneity in the presence or absence of the DNMT inhibitors decitabine (Dec) and 5-aza-cytidine (Aza), and the HDAC inhibitor vorinostat (Vor) (Figs. 4, D to H, and 5).

As shown schematically in Fig. 4D, cells were preincubated with Dec, Aza, Vor, or vehicle alone at sublethal doses for 2 days (~25% reduction in growth rate; see fig. S6A for dose-survival curves). They were then encapsulated and grown in CAGEs over a period of 6 days with continued drug treatment. At progressive time points, the clones were sampled, lysed in CAGEs, and then frozen before inC-RNA-seq. In total, we analyzed transcriptomes from 134,805 CAGEs representing single cells sampled at day 0 ($n = 37{,}902$) and expanded clones grown 2, 4, or 6 days ($n = 96{,}903$) in different treatment conditions. An

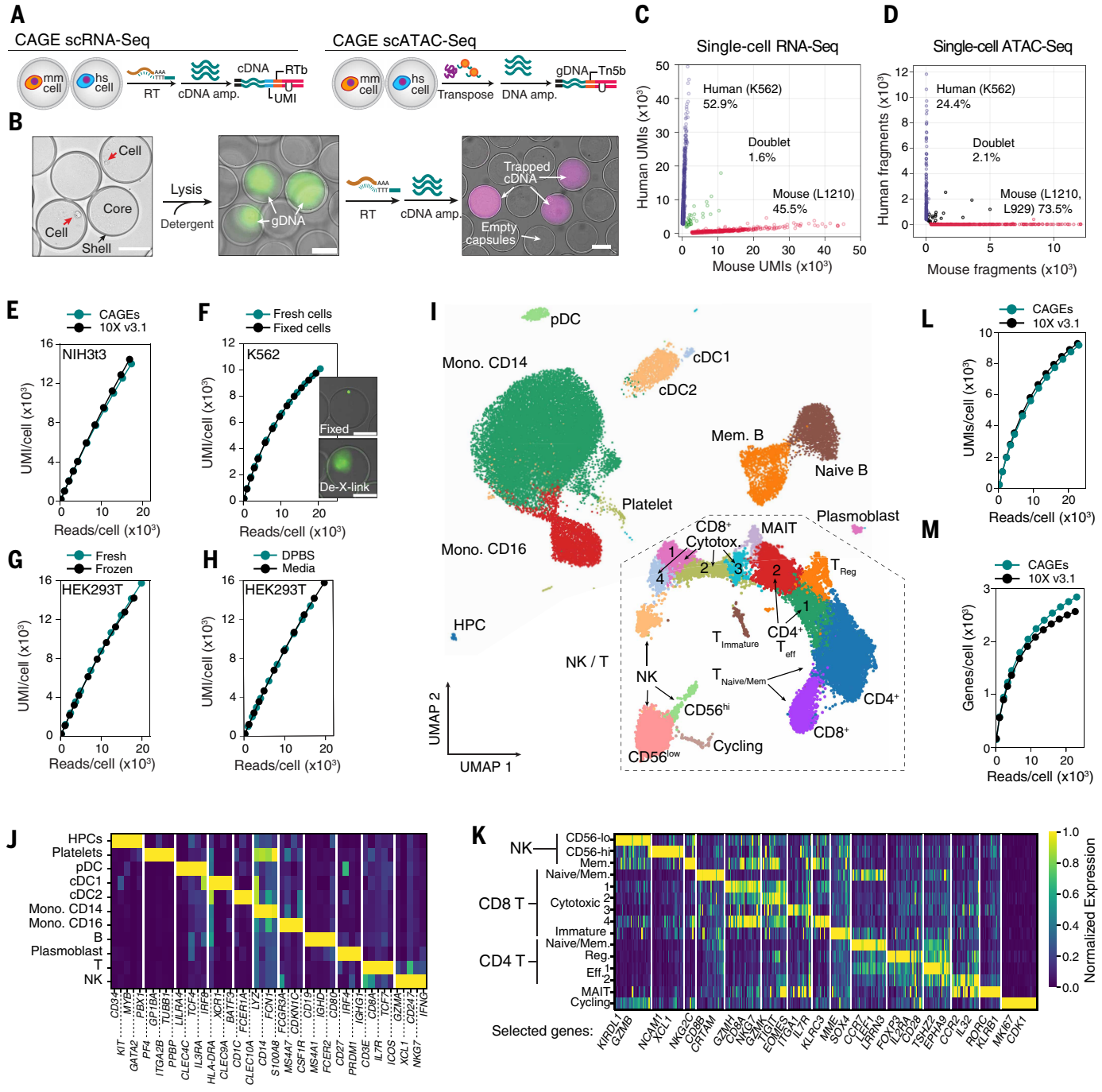


Fig. 3. CAGEs enable single-cell genomic assays with multistep processing. (**A**) Schematic of inC-RNA-seq and inC-ATAC-seq library preparation modalities. Additional barcodes and sequences are added during transcript [reverse-transcription barcode (RTb) and unique molecular identifier (UMI)] and accessible gDNA region (Tn5b) capture through RT and tagmentation steps. Captured and amplified dU-bearing molecules are subjected to split-pool barcoding and sequencing. mm cell, mus musculus cell; hs cell, homo sapiens cell; cDNA amp., complementary DNA amplicon. (**B**) Micrographs showing cells at progressive stages of scRNA-seq library preparation in CAGEs, from initial cell encapsulation through lysis, to generation of amplified cDNA libraries before in-CAGE barcoding. gDNA is stained with SYBR Safe dye; cDNA is stained with Cy5 fluorophores conjugated to the PCR amplification primers. (**C** and **D**) Plots showing the number of UMIs or DNA fragments corresponding to each CAGE mapping to human (*y* axis) and mouse (*x* axis) RNA-seq (left) or ATAC-seq (right) inC-seq libraries. The observed mixed species doublet rates are in line with the expectations: 2.1% for (C) with (loading of 0.083 cells per CAGE) and 1.9% for (D) (loading of 0.1 cells per CAGE). (**E**) UMI per reads plots after down-sampling reads from CAGE scRNA-seq libraries in comparison with publicly available 10x Genomics NextGem v3.1 data. (**F** to **H**) UMI per reads plots as shown in (E) for libraries generated from fresh or paraformaldehyde-fixed K562 cells (F), fresh or frozen HEK293T cells (G), or capsules made using DPBS or media (H). (**I**) A two-dimensional UMAP embedding of inC-RNA-seq data from PBMCs. MAIT, mucosal-associated invariant T cell; Mem. B, memory B cell; Mono. CD14, monocyte CD14; pDC, plasmacytoid dendritic cell; T_{reg}, regulatory T cell. (**J**) Heatmaps of gene expression showing coverage of low- and high-abundance cell type–specific genes. (**K**) High population structure of partitioning subsets of T and NK cells. (**L**) UMI per reads plots as shown in (F), now generated for PBMC data and comparing inC-seq with data from 10x Genomics GEMCode. (**M**) Comparison of number of genes detected per cell as a function of read depth between inC-seq and 10x Genomics NextGem v3.1.

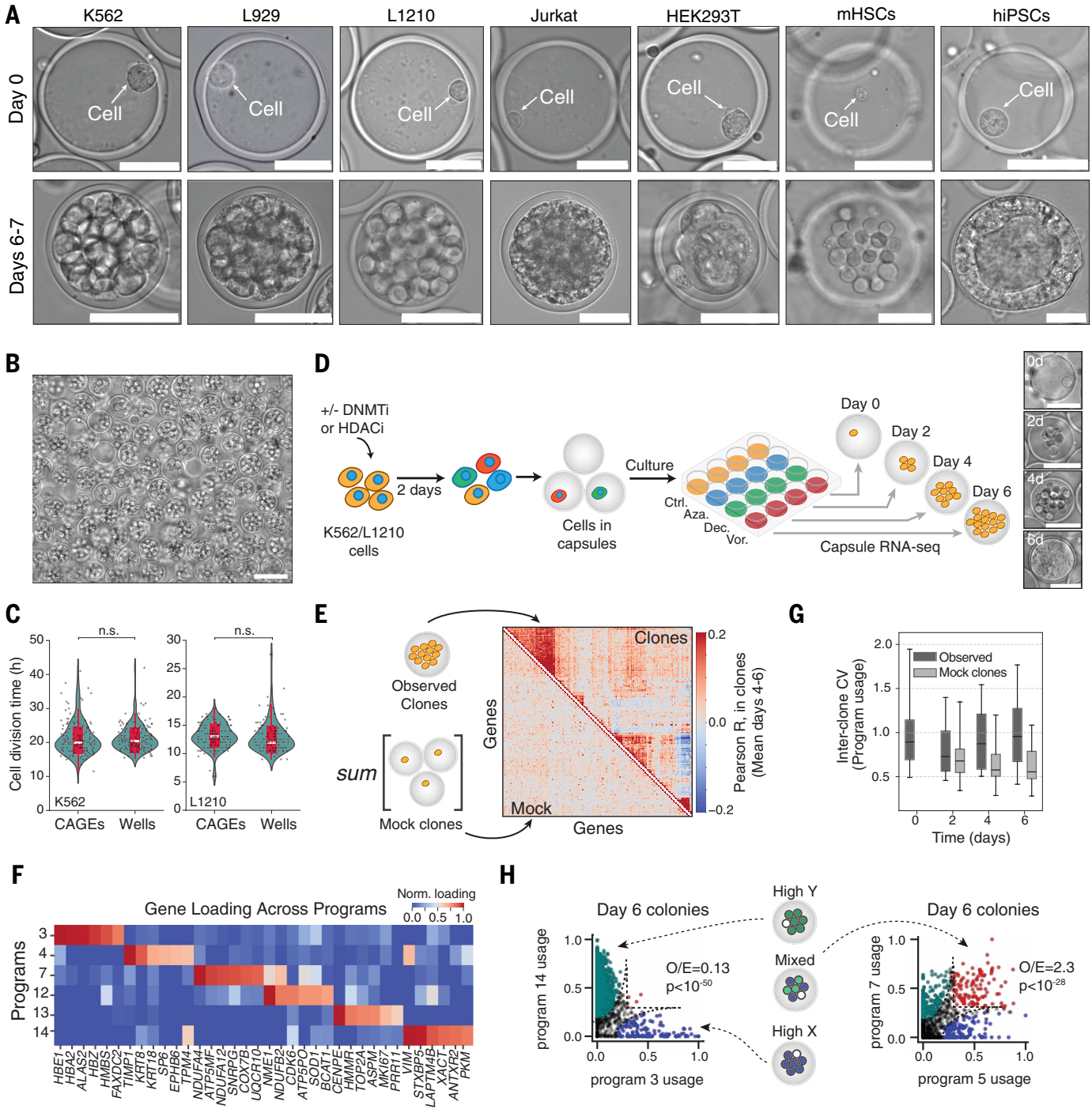


Fig. 4. Live-cell clonal expansion in CAGEs reveals persistent gene expression programs. (**A**) Micrographs showing growth in CAGEs of clones from cultured cell lines, primary mouse bone marrow hematopoietic stem cells (mHSCs), and human induced pluripotent stem cells (hiPSCs). Scale bars, 50 μm. (**B**) Micrograph of encapsulated colonies concurrently cultured. Scale bar, 100 μm. The image shows high-density K562 (polyclonal) colonies expanded in CAGEs. (**C**) Division times of cells cultured in wells or CAGEs. Each dot corresponds to a single cell-division time evaluated by time-lapse microscopy (K562: $n = 156$; L1210: $n = 127$). *t* test *P* value >0.8. (**D**) Experimental schema for identifying persistent gene expression programs in cancer cell lines upon treatment with vehicle (Ctrl), DNMT inhibitors [5-azacytidine (Aza); decitibine (Dec)], or an HDAC inhibitor [vorinostat (Vor)]. Micrographs show example capsules at different time points. (**E** to **H**) Evidence of persistent gene expression programs in K562 cells; see fig. S6 for L1210 cells. (E) Clustered interclonal gene-gene expression correlations in K562 cell control samples at days 4 to 6 show evidence of structured programs, which are absent when randomly combining cells sampled at day 0 into "mock" clones. (F) Top genes contributing to gene expression programs identified by NMF, identifying erythroid, cell cycle, and vimentin- and keratin-associated programs. Full program loadings found in table S3. (G) Box plots of variation in 15 NMF-derived programs between control clones, showing persistent heterogeneity over time compared with mock clones. (H) Some gene expression programs remain mutually exclusive after clonal expansion, as seen from low numbers of clones observed mixed NMF programs as compared with expectation [observed/expected (O/E) = $f_{XY}/f_X f_Y$; *P* value from Fisher's exact test].

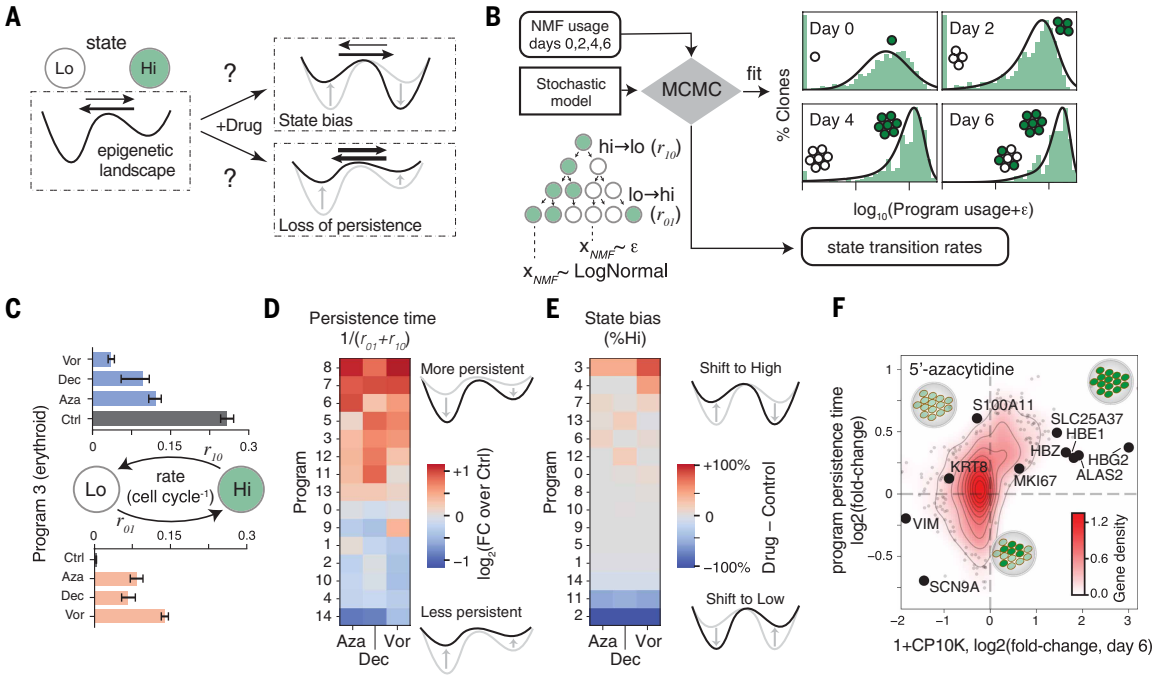


Fig. 5. Altered clonal memory upon treatment with DNMT or HDAC inhibitors. (**A**) A schematic of possible effects of drug treatment on epigenetic memory. In untreated controls, cells occupy high and low gene expression states for each observed program. Inhibitors of epigenetic modifying enzymes may bias the relative stability of states (top scenario), and may also reduce persistence (shallower wells, bottom scenario). From inC-RNA-seq on growing clones, one can infer state transition rates (arrows). (**B**) Schema for statistical inference of transition rates from clonal inC-RNA-seq data by fitting stochastic models of cell state transitions to observed NMF program usage over time. The model was used to infer rates of program induction (r_{01}) and loss (r_{10}) for each program in each treatment condition (see supplementary text 1). MCMC, Markov chain Monte Carlo. (**C**) Fitted transition rates exemplified for one program (program 3), with all drugs increasing persistence of the "high" state, while destabilizing the "low" state. (**D**) Changes in K562 gene expression program persistence times from the fitted dynamic transition rates, showing reduced clonal memory for some programs (blue) but not others (red). (**E**) Corresponding changes in steady-state bias (the fraction of cells in the active state), showing distinct changes from program persistence. (**F**) Changes in persistence are distinct from changes in mean expression for single genes, shown for one drug (Aza) in K562 cells.

equivalent analysis of this number of isolated colonies in wells would require 1010 96-well plates. In both K562 and L1210 clones, we identified genes with above Poisson CVs in gene expression (CV, or SD and mean; shown for untreated cells at days 4 and 6 in fig. S6, B and C). The most variable genes between clones defined modules of gene expression that persisted over time, as seen from clustering gene-gene correlations and comparing the results with those from mock clones generated by randomly combining single-cell transcriptomes (Fig. 4E and fig. S6D). Persistent clonal heterogeneity can also be appreciated by UMAP embedding of the clones (fig. S6E).

To identify persistent gene expression programs varying between clones, we decomposed their expression by non-negative matrix factorization (NMF). Approximately 12 to 15 programs explained variation in gene expression above random for both cell lines [eigenvalue crossover method (*5*); fig. S6F]. Several of these were readily identifiable from their gene loadings, including cell cycle–associated programs (K562 programs 12 and 13 and L1210 program 14); erythroid differentiation (K562 program 3); putative epithelial (*KRT8/18*-enriched) and mesenchymal (*VIM*-enriched) programs (K562 programs 4 and 14); and myeloid programs expressing the transcription factors *Irf8* and *Myc*, interleukin receptor *Il7r* (L1210 program 0), myeloid transcription factor *Klf6*, M-CSF (*Csf1*), and osteopontin (*Spp1*) (L1210 program 7) (Fig. 4F, fig. S6G, and table S4). These programs remained variable after clonal expansion beyond that expected from mock clones (Fig. 4G and fig. S6H). Several programs also showed mutually exclusive patterns of expression: For example, K562 day 6 colonies were either enriched in program 14 (*VIM*-hi) or program 3 (erythroid, *HBE1/HBA2/ALAS2*-hi) but not both ($P < 10^{-50}$, Fisher's exact test, adjusted to correct for false discovery). By contrast, mock colonies showed mixing of the two programs (Fig. 4H and fig. S6, I and J).

Next, we explored how Aza, Dec, or Vor treatments alter gene expression dynamics. Drug treatment that alters epigenetic inheritance could bias the fraction of cells that express a program, or alter the rate at which a program switches states, or both (Fig. 5A). To quantify these two effects, we introduced a statistical inference framework that fits the observed distribution of NMF program usages across clones over time to a stochastic model of state switching, parameterized by the persistence times of cells in a "low" state ($1/r_{01}$) and "high" state ($1/r_{10}$) for each program (Fig. 5, B and C, and supplementary text 1). From these fits we obtained changes in program persistence times [the relaxation timescale to steady state, $1/(r_{01} + r_{10})$] (Fig. 5D and fig. S7, A to C) and steady-state bias [the fraction of time that a program is active, $r_{01}/(r_{01} + r_{10})$] (Fig. 5E and fig. S7D). These analyses revealed that all three drugs triggered changes in program persistence and bias in K562 cells, with the dynamics of some programs slowing down (higher persistence) and others accelerating. In K562 cells, all three drugs triggered differentiation toward the erythroid state (program 3 high state), with increased persistence in the differentiated state (Fig. 5, C to E). The changes in persistence times were correlated but distinct from changes in mean expression of individual genes (Fig. 5F). In L1210 cells, the DNMT inhibitors (Aza and Dec) showed distinct responses from the HDAC inhibitor (fig. S7, C and D), but again the drugs led to both increases and decreases in state persistence times. Thus, contrary to what might be expected, the DNMT and HDAC inhibitors did not reduce clonal heterogeneity over several cell divisions. This observation is consistent with prior work that showed only modest changes in gene expression heritability after genetic

ablation of DNA methyltransferases (*36*). The specific nature of the effects of these drugs, and their failure to globally reduce persistence, may explain why they sensitize chemotherapy in some patients but not others (*37*). More generally, the approaches taken here for live-cell genomic analysis after perturbations and growth may be generalizable to other colony assays, including on gastruloids and tissue-derived organoids.

Discussion

We have established a versatile platform for high-throughput single-cell and single-colony genomic analysis through the development and use of capsules with amphiphilic gel envelopes—CAGEs. The small compartments are biocompatible, quick to produce, and have tunable chemistry. Alternative compositions can be evaluated through biophysical assays to define new permeability, mechanical, adhesive, or other functional properties. The approaches reported here for barcoding DNA libraries in capsules (inC-seq) and to study living cells can be adapted to a range of problems and may be further combined with sorting, fixing, and staining steps, each of which is demonstrated here.

CAGEs also have some limitations. First, when CAGEs are used for single-cell assays, many are empty and increase the volume of subsequent reactions. This problem is alleviated by performing experiments in which multiple cells are encapsulated or by enriching for CAGEs containing cells. Second, although CAGEs are compatible with most molecular biology reactions, they may require slight changes in conditions to maximize efficiency and yield. The inC-RNA-seq workflow described in this work, for example, was optimized by changing reagent concentrations and reaction times, whereas inC-ATAC-seq is demonstrated but not yet optimized. Lastly, the F127DA CAGEs described in this work are not suitable for live-cell recovery because they can only be solubilized by using harsh reaction conditions. Alternative CAGEs may be amenable to enzymatic solubilization.

Overall, CAGEs present a powerful tool for live-cell functional and genomic assays. They will likely enable the profiling of difficult-to-process samples and may prove useful in studying perturbation outcomes that depend on cell-cell interactions, such as developmental or immune-cell interactions, and for performing colony-forming assays or studies of organoids at scale. Capsules may also allow the evaluation of libraries of signaling molecules or may be used as vehicles for clonal evolution or selection followed by genomic profiling. These in turn may provide functional information that could support the development of predictive, data-driven models of cellular dynamics and interactions.

REFERENCES AND NOTES

1. K. Vandereyken, A. Sifrim, B. Thienpont, T. Voet, *Nat. Rev. Genet.* **24**, 494–515 (2023).
2. A. Baysoy, Z. Bai, R. Satija, R. Fan, *Nat. Rev. Mol. Cell Biol.* **24**, 695–713 (2023).
3. M. Hagemann-Jensen *et al.*, *Nat. Biotechnol.* **38**, 708–714 (2020).
4. M. Hagemann-Jensen, C. Ziegenhain, R. Sandberg, *Nat. Biotechnol.* **40**, 1452–1457 (2022).
5. A. M. Klein *et al.*, *Cell* **161**, 1187–1201 (2015).
6. T. M. Gierahn *et al.*, *Nat. Methods* **14**, 395–398 (2017).
7. I. C. Clark *et al.*, *Nat. Biotechnol.* **41**, 1557–1566 (2023).
8. J. Cao *et al.*, *Science* **361**, 1380–1385 (2018).
9. A. B. Rosenberg *et al.*, *Science* **360**, 176–182 (2018).
10. J. de Rutte *et al.*, *ACS Nano* **16**, 7242–7257 (2022).
11. S. Udani *et al.*, *Nat. Nanotechnol.* **19**, 354–363 (2024).
12. J. L. McFaline-Figueroa *et al.*, *Cell Genom.* **4**, 100487 (2024).
13. G. H. T. Yeo *et al.*, *Cell Stem Cell* **26**, 938–950.e6 (2020).
14. G. Leonaviciene, L. Mazutis, *Nucleic Acids Res.* **51**, e2 (2023).
15. G. Leonaviciene, K. Leonavicius, R. Meskys, L. Mazutis, *Lab Chip* **20**, 4052–4062 (2020).
16. D. Baronas *et al.*, *Science* **2025** eady7227 (2025).
17. T. Y. Lee, T. M. Choi, T. S. Shim, R. A. M. Frijns, S.-H. Kim, *Lab Chip* **16**, 3415–3440 (2016).
18. S. Ma *et al.*, *Small* **8**, 2356–2360 (2012).
19. M. Yanagisawa, S. Nigorikawa, T. Sakaue, K. Fujiwara, M. Tokita, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15894–15899 (2014).
20. D. Attwood, J. Collett, C. Tait, *Int. J. Pharm.* **26**, 25–33 (1985).
21. M. Di Biase *et al.*, *Soft Matter* **7**, 4928 (2011).
22. A. Y. L. Sim, J. Lipfert, D. Herschlag, S. Doniach, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **86**, 021901 (2012).
23. Y. Lu, B. Weers, N. C. Stellwagen, *Biopolymers* **61**, 261–275 (2001-2002).
24. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, *Curr. Protoc. Mol. Biol.* **109**, 29.1, 9 (2015).
25. T. Scully, A. Klein, A mannitol-based buffer improves single-cell RNA sequencing of high-salt marine cells. bioRxiv 2023.04.26.538465 [Preprint] (2023). https://doi.org/10.1101/2023.04.26.538465.
26. M. De Simone *et al.*, Comparative Analysis of Commercial Single-Cell RNA Sequencing Technologies. bioRxiv 2024.06.18.599579 [Preprint] (2024). https://doi.org/10.1101/2024.06.18.599579.
27. B. Xia, K. Krutkramelis, J. Oakey, *Biomacromolecules* **17**, 2459–2465 (2016).
28. S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, P. K. Sorger, *Nature* **459**, 428–432 (2009).
29. S. M. Shaffer *et al.*, *Nature* **546**, 431–435 (2017).
30. A. Gnyszka, Z. Jastrzebski, S. Flis, *Anticancer Res.* **33**, 2989–2996 (2013).
31. J. E. Bolden, M. J. Peart, R. W. Johnstone, *Nat. Rev. Drug Discov.* **5**, 769–784 (2006).
32. J. Strauss, W. D. Figg, *Anticancer Res.* **36**, 1–4 (2016).
33. B. T. Oronsky *et al.*, *Front. Oncol.* **5**, 134 (2015).
34. M. Szyf, *Annu. Rev. Pharmacol. Toxicol.* **49**, 243–263 (2009).
35. S. M. Shaffer *et al.*, *Cell* **182**, 947–959.e17 (2020).
36. Z. Meir, Z. Mukamel, E. Chomsky, A. Lifshitz, A. Tanay, *Nat. Genet.* **52**, 709–718 (2020).
37. K. Nepali, J.-P. Liou, *J. Biomed. Sci.* **28**, 27 (2021).
38. I. Mazelis, A. M. Klein, AllonKleinLab/Mazelis2025_CAGEs: Paper release, version v0.91, Zenodo (2025); https://doi.org/10.5281/zenodo.17373720.

ACKNOWLEDGMENTS

We thank L. Mazutis, D. Baronas, and T. Mitchison for ongoing discussions and feedback during this project; J. Oakey for guidance of ROS generation during photopolymerization; the labs of O. Pourquié and F. Camargo for providing iPSC and HSC cells; S. McGeary for providing K562 and L929 cells; Y. Chen for providing L1210 cells; the lab of M. Greenberg for access to their Illumina Sequencer; J. A. Nelson at the Bauer Core Facility at Harvard University for FACS work coordination; and the Single Cell Core at HMS for supporting work by T.L.T. and A.K. The cryo-SEM work was performed in part at the Harvard University Center for Nanoscale Systems (CNS), a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the NSF under NSF award no. ECCS-2025158. We extend our heartfelt gratitude to A. Graham at the Harvard CNS core, whose expertise was instrumental in bringing this cryo-SEM work to fruition. We thank all members of the Klein lab for critical reading and feedback on the manuscript. **Funding:** This work was supported by NIH grants R21HG012771 and R33CA278392, and by an Edward Mallinckrodt Jr Scholar Award to A.M.K. Pilot work on the genomic assays was supported by an HMS Q-FASTR Pilot Grant and pilot work solving viability in live-cell assays by an HMS Blavatnik Biomedical Accelerator Pilot Grant. The Harvard University CNS is a member of the NNCI, which is supported by the NSF under NSF award no. ECCS-2025158. **Author contributions:** I.M. developed the capsules and carried out all experiments except FACS and Cryo-SEM; H.S. supported diffusion assays and optimizing cell culture conditions in capsules. T.L.T. carried out flow sorting in capsules; A.K. carried out SEM imaging with support of the CNS core facility; I.M. and A.M.K. conceived and designed experiments and carried out data analysis; A.M.K. supervised the work; and I.M. and A.M.K. wrote the manuscript. All authors provided feedback on the manuscript. **Competing interests:** A.M.K. is a cofounder of Somite Therapeutics. I.M. and A.M.K. are inventors on patent application PCT/US2023/029364 filed by Harvard University. **Data, code, and materials availability:** Sequencing data are available at Gene Expression Omnibus (GSE306708) and Sequence Read Archive (PRJNA1311090). Python code for capsule segmentation and pipeline for scRNA-seq read preprocessing are archived at Zenodo (*38*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ady7209
Materials and Methods; Supplementary Text; Figs. S1 to S7; Tables S1 to S5; References (*39*–*46*); MDAR Reproducibility Checklist; Movie S1

Submitted 2 May 2025; accepted 5 December 2025; published online 18 December 2025
10.1126/science.ady7209

Multistep genomics on single cells and live cultures in subnanoliter capsules

Ignas Mazelis, Haoxiang Sun, Arpita Kulkarni, Theresa L. Torre, and Allon M. Klein

Science **391** (6790), . DOI: 10.1126/science.ady7209

Editor's summary

Profiling single cells across multiple modalities requires complex, multistep molecular biology workflows that are difficult to scale. Two studies developed semipermeable capsule (SPC) technology in which microscopic liquid vessels are enclosed by a thin, biodegradable shell that efficiently retains nucleic acids of encapsulated cells while remaining permeable to enzymes, oligonucleotides, and other reagents. Mazelis *et al*. implemented a barcoding method to enable single-capsule readouts and then used these barcodes to study how gene expression patterns persist as cells divide. Baronas *et al*. implemented a single-cell RNA-sequencing approach that enables accurate and comprehensive capture of diverse cell types in circulating cells of acute myeloid leukemia patients, uncovering distinct leukemic cell phenotypes. SPCs offer a versatile and scalable platform for single-cell omics assays, including RNA and genome sequencing, nucleic acid–guided cell sorting, and long-term cell cultivation. —Di Jiang